Discovery and biochemical characterization of RNA interference in budding yeast

by

David E. Weinberg

B.S./M.S. Molecular Biophysics & Biochemistry
Yale University, 2007

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2013

Signature of Author: ______________________________
David E. Weinberg
Department of Biology
May 21, 2013

Certified by: ______________________________
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by: ______________________________
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

Discovery and biochemical characterization of RNA interference in budding yeast

by

David E. Weinberg

Submitted to the Department of Biology on May 21, 2013
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

RNA interference (RNAi) is a eukaryotic pathway for the post-transcriptional regulation of gene expression. In the simplest form of RNAi, a double-stranded RNA (dsRNA) trigger is converted into small-RNA duplexes by the Dicer enzyme. These duplexes are then loaded into the effector protein Argonaute to guide the cleavage of target transcripts. RNAi and related RNA-silencing pathways are found in plants, animals, fungi, and protists, suggesting origins in an early eukaryotic ancestor and selective pressures to maintain the pathway. A prominent exception to this widespread conservation of RNAi is the budding yeast *Saccharomyces cerevisiae*, which lacks homologs of Dicer and Argonaute. Indeed, RNAi had been presumed lost in all budding yeasts.

Motivated by the presence of Argonaute homologs in some budding-yeast species, we examined whether these species contain a functional RNAi pathway. High-throughput sequencing led to the identification of endogenous small RNAs that are generated by a novel Dicer enzyme. In *Saccharomyces castellii*, these Argonaute-bound small RNAs serve as guides to repress mRNA targets, which are predominantly repetitive elements. RNAi can be restored to *S. cerevisiae* by introducing the genes encoding *S. castellii* Dicer and Argonaute, and the reconstituted pathway silences endogenous transposons.

Budding-yeast Dicer has a different domain architecture than canonical Dicer yet generates siRNAs of a similar length. In contrast to canonical Dicer, which successively removes small-RNA duplexes from the dsRNA termini, budding-yeast Dicer molecules bind cooperatively to the interior of dsRNA substrates, with the distance between adjacent active sites determining product length. These distinct mechanisms impart corresponding substrate preferences and product characteristics that are important for Dicer function.

Structural studies of budding-yeast Argonaute yielded a crystal structure of the functional Argonaute–guide complex. Eukaryotic Argonaute makes extensive sequence-independent interactions with the guide RNA to maintain the seed region in a helical conformation with the base edges accessible for target binding. An invariant glutamate residue, which is only positioned in the catalytic pocket after guide-RNA binding, constitutes the previously missing component of a ribonuclease H–like active site.

Thesis Supervisor: David P. Bartel
Title: Professor of Biology

Acknowledgements

First and foremost, I thank Dave Bartel for the opportunity and privilege to work in his lab. He has been an amazing mentor with respect to both bench science and career advice, and his enthusiasm and dedication to science are infectious.

I thank my undergraduate mentor, Sandy Wolin, who took me into her lab as a naïve freshman and taught me how to think critically about science. I also thank my biochemistry professors at Yale, Joan Steitz and Scott Strobel, who unknowingly convinced me that RNA is the most interesting biological macromolecule.

I am indebted to my structural-biology collaborators, Kotaro Nakanishi and Dinshaw Patel, for their invaluable contributions to my thesis research. It has been a privilege to work with such dedicated and creative scientists. I am also grateful to Gerry Fink and members of his lab, particular Doug Bernstein and Valmik Vyas, for welcoming me into their lab meetings and for always being so willing to share their expertise.

I thank my thesis committee members, Phil Sharp and Steve Bell, who have followed me patiently through my graduate career and always provided lucid advice and thoughtful insights. I also thank Bob Sauer, Frank Solomon, Iain Cheeseman, David Sabatini, and Wendy Gilbert for many fruitful discussions.

I am grateful to every member of the Bartel lab with whom I overlapped (Vikram, Vincent, Daehyun, Rosaria, Alvin, Grace, Anna, Stephen, Wenwen, Robin, David G., Matt, Andrew, Huili, Junjie, Katrin, Sue-Jean, Calvin, Wendy, Jinkuk, Ben, David K., Gina, Mike, Christine, Sean, Jinwu, Michael, Laura, Olivia, Graham, David S., Lori, Chanseok, Lena, Noah, Asia, Alex, Igor, Kathy, and Muhammed) for sitting through countless lab meetings and practice talks, for reading endless drafts of research proposals, and for many hours of discussions and amusement. I am especially thankful to my past and present baymates, Huili and Matt, and my fellow inhabitants of room 623, Jinkuk, Sue-Jean, and Alex, for putting up with my silly antics over the years.

I thank the members of the yeast subgroup (Anna, Kathy, Lori, David G., David K., Matt, and Katrin) for productive discussions and for creating a collaborative atmosphere. I am particularly grateful to Anna, who mentored me when I was a rotation student and allowed me to join the budding-yeast project, and Kathy for her important contributions that made some of my experiments possible.

I thank my classmates in Biograd2007, especially Zach, Keren, Liron, Emily, and Lauren, who have helped to make the last five years fun. I am grateful to my fellow inhabitants of 10 Michael Way, Mike, Irving, and Dann, for keeping me sane and providing constant entertainment at home.

Last but not least, I thank my family. I thank my parents for ensuring that I received the best education possible and for being so supportive of my career path. I thank my brothers for all of the decidedly non-scientific experiences that we’ve shared.

Table of Contents

Chapter 1

Introduction

Part 1: A brief history of RNAi

Discovery of RNA interference

What later came to be known as a manifestation of RNA interference, or RNAi, began as a curious observation in plants. In efforts to study the pigmentation of petunia petals, transgenes encoding components of the pigment-biosynthesis pathway were introduced into petunias and unexpectedly led to silencing of both the transgene and the corresponding endogenous gene, giving rise to unpigmented petunias (Napoli et al., 1990; van der Krol et al., 1990). This ability for the transgene to be coordinately silenced with the endogenous gene was referred to as "co-suppression" (Napoli et al., 1990). A similar phenomenon was soon observed in the filamentous fungus *Neurospora crassa*[1] and termed "quelling" (Romano and Macino, 1992). Notably, both co-suppression and quelling were reversible, and transformants that reverted to wild-type or intermediate phenotypes showed increased levels of both the transgenic and endogenous transcripts, consistent with the coordinated changes observed in the silenced state (Napoli et al., 1990; Romano and Macino, 1992).

The first indication that these silencing mechanisms may not just be a peculiarity of the plant and fungal kingdoms came only a few years later. Antisense inhibition had been recently introduced as a technique for silencing genes in the nematode *Caenorhabditis elegans* by transgenic expression of an antisense transcript (Fire et al., 1991). Although the molecular mechanism of antisense inhibition was unknown, it was noted that even control constructs designed to express the corresponding sense transcript led to silencing of the endogenous gene in

---

[1] Also revealed as a loss-of-pigmentation phenotype

some cases. Despite the striking similarity of this phenomenon to co-suppression in plants (Napoli et al., 1990; van der Krol et al., 1990), the observed sense inhibition in worms was initially interpreted as an artifact of spurious generation of antisense RNAs from the sense constructs, which caused antisense inhibition (Fire et al., 1991). However, subsequent efforts to use antisense inhibition via direct injection of RNA into the germline also revealed an inhibitory effect of the sense RNA (Guo and Kemphues, 1995), a result that directly implicated the sense transcript in silencing.

The unification of these peculiar observations came with the discovery that double-stranded RNA (dsRNA) was a far more potent trigger of silencing than either sense or antisense RNAs alone when injected into worms (Fire et al., 1998). Thus, it was reasoned that sense inhibition and antisense inhibition were related to the unintentional (and quite fortuitous) production of low levels of dsRNA, which was the initiator molecule of RNAi. Silencing by dsRNA was exceptionally specific to the targeted gene and extremely effective, in most cases yielding a phenotype that mirrored the known null-mutant phenotype. The mechanism of RNAi appeared to be post-transcriptional down-regulation of target RNA transcripts, as evidenced by an inability to provoke silencing using dsRNA targeting promoter regions or introns and by a dramatic decrease in the levels of the endogenous mRNA transcripts.

Small RNAs and Dicer

Having established dsRNA as the trigger for RNAi (Fire et al., 1998), efforts focused on the molecular basis for target specificity. The sequence specificity and post-transcriptional nature of RNAi immediately suggested the formation of a duplex between the target mRNA and antisense strand of the dsRNA trigger (Palauqui and Balzergue, 1999). In the absence of reports that such long antisense RNAs were specifically present in organisms exhibiting RNAi, attention turned to smaller RNAs that may have been missed[2] by traditional analyses. The first indication that such small RNAs might be involved in RNAi came with the discovery of a population of ~25-nucleotide (nt) RNAs in plants that corresponded to the silenced gene and were only present in plants exhibiting post-transcriptional gene silencing (PTGS[3]) (Hamilton and Baulcombe, 1999). Despite the fact that these RNAs were significantly smaller than the RNA strands comprising the initial dsRNA triggers, it was appreciated that the small RNAs were still sufficiently long to convey sequence specificity. However, it was not known whether these small RNAs were derived directly from processing of the dsRNA trigger or were a byproduct of a downstream step in the pathway.

Biochemical analysis of the RNAi pathway was enabled by the development of a cell-free system derived from *Drosophila* embryos that recapitulated dsRNA-induced target-RNA degradation (Tuschl et al., 1999). In this *in vitro* system, addition of dsRNA corresponding to a luciferase reporter gene led to a reduction in the production of luciferase protein and a corresponding disappearance of the luciferase mRNA transcript, indicative of mRNA degradation. Notably, preincubation of the dsRNA in lysate prior to addition of target RNA enhanced target-RNA degradation, suggesting that covalent modification of the dsRNA might be required for an activation step prior to target-RNA degradation. Indeed, using radiolabeled substrates in this system it was possible to follow the fate of each RNA molecule and thereby demonstrate that the dsRNA was processed into a population of 21–23-nt species that corresponded to both strands (Zamore et al., 2000). This *in vitro* activity that cleaved dsRNA into small RNA fragments implicated a member of the ribonuclease III (RNase III) family of

---

[2] Or largely ignored as degradation products rather than functional molecules

[3] A specific manifestation of co-supression that acts by a post-transcriptional mechanism

dsRNA-specific endoribonucleases (Bass, 2000). Consistent with processing by an RNase III enzyme, the dsRNA-derived small-RNA products contain the hallmarks of RNase III–mediated cleavage: 5′ phosphate and 3′ hydroxyl termini on each strand, with the opposite strands offset by a 2-nt 3′ overhang (Elbashir et al., 2001a).

Analysis of the genome sequences of *Drosophila melanogaster* and *Caenorhabditis elegans* identified a class of RNase III enzymes that contained two RNase III domains and an amino-terminal helicase domain (Bernstein et al., 2001), which on the basis of this domain structure had previously been proposed to be the dsRNA-processing enzyme (Bass, 2000; Cerutti et al., 2000). Biochemical analysis of one of the *Drosophila* homologs demonstrated the ability to convert long dsRNA into ~22-nt products in a sequence-independent manner, an activity for which the enzyme was named Dicer (Bernstein et al., 2001). The involvement of Dicer in RNAi *in vivo* was further supported by genetic experiments demonstrating that interfering with Dicer function (either by dsRNA-mediated knockdown of the gene in *Drosophila* S2 cells or by mutant analysis in *C. elegans*) led to defects in RNAi (Bernstein et al., 2001; Knight and Bass, 2001).

With a defined and essential molecular function in the RNAi pathway, Dicer was harnessed as an anchor on which to gain phylogenetic insights into the distribution of the RNAi pathway. Consequently, with the discovery of Dicer came an immediate appreciation for its widespread evolutionary conservation, with homologs found in plants, animals, and fungi (Bernstein et al., 2001). Notably, the small-RNA-generating activity of the human Dicer homolog (Bernstein et al., 2001) coupled with observations of dsRNA-induced gene silencing in mouse embryos (Wianny and Zernicka-Goetz, 2000) provided the earliest indications that an RNAi pathway may even function in mammals.

Guide RNAs

Following the discovery that the dsRNA trigger was processed by Dicer into small RNAs, it remained to be determined if these small RNAs were ultimately guiding target-RNA degradation. In *Drosophila* lysate, degradation of a radiolabeled target RNA by RNAi generated stable 5′ cleavage products that were restricted to the region corresponding to the dsRNA (Zamore et al., 2000). Precise mapping of the cleavage positions indicated that the target RNA was cleaved at 21–23-nt intervals, the same interval that spanned Dicer cleavage products. Together, these observations suggested that the small RNAs acted as the specificity determinant that directly guided target-RNA cleavage. Consistent with this notion, biochemical purification of the nuclease responsible for target cleavage identified a population of small RNAs that copurified with the activity (Hammond et al., 2000). In an *in vivo* RNAi system triggered by dsRNA, the kinetics of the appearance small RNAs mirrored the kinetics of target-RNA degradation, consistent with a causal link (Yang et al., 2000). Moreover, when S2 cells were transfected with radiolabeled dsRNAs, the nuclease activity responsible for RNAi copurified with radiolabeled small RNAs (Bernstein et al., 2001). Together, these observations indicated that the small RNAs generated by Dicer are the specificity determinants that guide RNA cleavage.

The notion that dsRNA is only required to generate the functional small RNA guides was ultimately proven by eliciting RNAi using synthetic small-RNA duplexes (Elbashir et al., 2001a). These duplexes, which contained 2-nt 3′ overhangs to mimic the products of Dicer cleavage, were named short interfering RNAs (siRNAs). Only the guide strand of the siRNA duplex remains associated with the active nuclease complex, indicating that the siRNA duplex is unwound and the other strand (designated the passenger strand) is discarded during the assembly process (Martinez et al., 2002). Indeed, synthetic single-stranded antisense RNAs can even guide target-RNA cleavage, indicating that only one strand of the siRNA duplex is required for the cleavage step of RNAi (Martinez et al., 2002; Schwarz et al., 2002).

RISC, Argonaute, and 'Slicer'

Despite rapid advances in understanding the relationship between the dsRNA trigger, siRNA effectors, and Dicer enzyme, the identity of the nuclease that catalyzes target-RNA cleavage remained unknown. It had originally been proposed that the same RNase III–family enzyme might cleave both the dsRNA trigger and the duplex formed between the guide and target RNAs (Bass, 2000). Although consistent with the double-stranded nature of the guide–target duplex, an RNase III enzyme would cleave both the target and guide RNAs, which would presumably render the complex a single-turnover enzyme. Alternatively, target-RNA cleavage might be mediated by a different nuclease that cleaves only the target strand of the guide–target duplex. Indeed, fractionation experiments indicated that Dicer and target-RNA cleavage activities can be biochemically separated, suggesting that a separate nuclease is responsible for target-RNA cleavage (Bernstein et al., 2001). This target-RNA nuclease was termed the RNA-induced silencing complex (RISC).

RISC activity was present in lysates prepared from S2 cells that had been transfected with dsRNA and had specificity corresponding to the dsRNA (Hammond et al., 2000). Biochemical purification of RISC followed by identification of constituent proteins by mass spectrometry identified a member of the Argonaute family of proteins (Hammond et al., 2001). Although this was the first biochemical evidence that Argonaute proteins were a component of RISC, Argonaute family members had been previously implicated in RNAi based on genetic studies in

C. elegans, *Neurospora*, and *Arabidopsis thaliana*[4] (Cogoni and Macino, 1997; Tabara et al., 1999; Catalanotto et al., 2000; Fagard et al., 2000). Despite this wealth of genetic and biochemical evidence that converged onto the Argonaute family, the precise mechanism by which Argonaute functioned as a component of RISC was not immediately obvious. Although it was recognized that Argonaute proteins contained PAZ and PIWI domains (Cerutti et al., 2000), at the time the functions of both domains were unknown and so the biochemical activity of Argonaute remained a mystery (Hammond et al., 2001).

Having dismissed Argonaute as unlikely to be the RISC nuclease, efforts focused on characterizing additional RISC components in order to identify the 'Slicer' within RISC (Hutvagner and Zamore, 2002). This resulted in the identification of two putative RNA-binding proteins (the *Drosophila* homolog of the fragile X mental retardation protein, dFXR; and Vasa intronic gene, or Vig) as well as the micrococcal nuclease homolog Tudor-SN (Caudy et al., 2002; Caudy et al., 2003). As a bona fide nuclease, Tudor-SN was an attractive candidate to fulfill the role of 'Slicer'. However, its lack of sequence specificity, generation of products bearing 3′-phosphate termini, and ability to cleave DNA in addition to RNA were inconsistent with the known biochemical properties of RISC (Caudy et al., 2003).

More detailed biochemical analysis of RISC-catalyzed degradation shed light on the properties of the enigmatic 'Slicer'. Cleavage occurs precisely at the phosphodiester bond positioned opposite nucleotides 10–11 of the guide RNA, with the 5′ end of the guide determining the cleavage position (Elbashir et al., 2001b). The cleavage reaction is a magnesium-dependent hydrolysis reaction that results in the generation of a 5′-monophosphate terminus and 3′-hydroxyl terminus (Martinez and Tuschl, 2004; Schwarz et al., 2004). Although

---

[4] The first Argonaute gene was identified in *Arabidopsis* and named for the mutant phenotype, which includes abnormal flowers with a squid-like appearance (Bohmert et al., 1998).

the products of the cleavage reaction are generally unstable, both the 5′ and 3′ cleavage products can be observed in ribonuclease-inhibited lysates, thus confirming that RISC functions as an endonuclease (Schwarz et al., 2004). Consistent with the guide RNA remaining intact after target cleavage, RISC behaves as a multiple-turnover enzyme; the rate-determining step *in vitro* is product release (Haley and Zamore, 2004). These properties of RISC, which contrasted with the known activities of Tudor-SN homologs, provided further evidence that Tudor-SN is not ‘Slicer’ (Schwarz et al., 2004).

Argonaute as ‘Slicer’

With Argonaute having emerged as a conserved component of RISC, structural studies of Argonaute were undertaken in order to understand how it contributes to RISC activity. Early efforts focused on bacterial Argonautes, which have the same domain structure as their eukaryotic counterparts. The crystal structure of the Argonaute protein from *Pyrococcus furiosus* revealed a bilobal architecture, with one lobe composed of the N and PAZ domains and the other lobe composed of the MID and PIWI domains (Song et al., 2004). A similar domain arrangement was observed in the structures of *Aquifex aeolicus* Argonaute and later *Thermus thermophilus* Argonaute, indicating that the bilobal architecture is a conserved feature of Argonaute proteins (Yuan et al., 2005; Wang et al., 2008b). Strikingly, the PIWI domain adopts an RNase H–like fold containing an active site comprised of highly conserved carboxylates, which strongly implicated Argonaute as the ‘Slicer’ (Parker et al., 2004; Song et al., 2004; Yuan et al., 2005). Furthermore, ‘Slicer’ cleavage products were known to have the same chemical features as RNase H cleavage products and both enzymes require magnesium ions for catalysis, which provided further support for the notion that the RNase H–like PIWI domain harbors ‘Slicer’ activity.

Biochemical support for the role of Argonaute as the catalytic component of RISC came from analysis of human Argonaute-2 (*Hs*AGO2), which demonstrated that variants harboring mutations in the putative catalytic residues were unable to assemble into active RISC either *in vitro* or *in vivo* (Liu et al., 2004). RISC activity was also purified to homogeneity from *Drosophila* S2 cells and found to only contain Argonaute-2 (Rand et al., 2004). Definitive proof that Argonaute and a guide RNA alone are sufficient to form RISC was enabled by the expression and purification of *Hs*AGO2 from *Escherichia coli*, which lacks an endogenous RNAi pathway. When combined with a single-stranded guide RNA, purified *Hs*AGO2 formed a complex that guided target-RNA cleavage with the features expected of RISC (Rivas et al., 2005). Thus, the cleavage activity of RISC resides within Argonaute itself.

Part 2: The RNAi machinery

With the recognition that Argonaute is the catalytic component of RISC, a basic outline of the RNAi pathway emerged (Figure 1). Long dsRNA is processed by Dicer into siRNA duplexes, which are loaded into a member of the Argonaute family of proteins. The resulting RISC, comprising Argonaute and a single-stranded guide RNA, binds target RNAs that are complementary to the guide RNA, and the enzymatic activity of Argonaute cleaves the target. In this simple scheme, the core RNAi machinery consists of just Dicer, Argonaute, and small RNAs.

Beyond Dicer and Argonaute: Additional RNAi proteins

In addition to this core machinery, additional protein cofactors facilitate RNAi *in vivo* (Figure 1). Indeed, although purified Argonaute and a single-stranded guide RNA assemble into RISC *in vitro*, purified Argonaute and an siRNA duplex do not form RISC *in vitro*, indicating that other proteins contribute to RISC formation (Liu et al., 2004; Rivas et al., 2005). Following generation by Dicer, siRNA duplexes are asymmetrically incorporated into Argonaute in a process referred to as RISC loading (Khvorova et al., 2003; Liu et al., 2003; Schwarz et al., 2003; Tomari et al., 2004a; Tomari et al., 2004b). A direct interaction between the PIWI domain of Argonaute and an RNase III domain of Dicer has been proposed to physically couple small-RNA production with RISC loading (Tahbaz et al., 2004). Moreover, low-resolution structural data suggests that RISC loading may proceed via an intermediate in which the Dicer PAZ domain and Argonaute PAZ domain simultaneously bind 2-nt 3′ overhangs at opposite ends of a small-RNA duplex (Wang et al., 2009a). Despite this proposed physical coupling, metazoan Dicer enzymes are unable to directly transfer a small-RNA duplex product into Argonaute. Instead they form complexes with double-stranded RNA-binding domain (dsRBD)–containing cofactors (i.e., TRBP in *Homo sapiens*, R2D2 in *Drosophila*, and HYL1 in *Arabidopsis*) that promote RISC assembly (Liu et al., 2003; Vazquez et al., 2004a; Chendrimada et al., 2005) (Figure 1). The ternary complex comprising Dicer, a dsRBD-containing protein, and Argonaute is referred to as the RISC-loading complex (RLC). In *S. pombe*, the ARC complex may play an analogous role in the loading of duplex siRNAs into Ago1 (Buker et al., 2007).

In addition to promoting the transfer of siRNA duplexes into Argonaute, the dsRBD-containing cofactors can function as asymmetry sensors that dictate which strand of the duplex becomes the guide RNA (Tomari and Zamore, 2005). Because each strand of the duplex has a unique targeting specificity, guide-strand selection determines which target RNAs will be silenced. Some small-RNA duplexes are incorporated into Argonaute based on the thermodynamic asymmetry of the duplex, with the less stable 5′ terminus being selected as the guide strand (Khvorova et al., 2003; Schwarz et al., 2003). The dsRBD cofactor is recruited to the more stable end and Dicer binds to the less stable end, thereby directing guide-strand selection for the duplex (Tomari et al., 2004b). In some cases Dicer can release a duplex product before recapturing it in its heterodimeric complex, rather than loading it directly into RISC (Vazquez et al., 2004b; Tomari and Zamore, 2005). This release–rebind mechanism enables sensing the thermodynamic asymmetry of the RNA duplex independently of the direction of Dicer processing.

Early *in vitro* studies of RNAi revealed a dependence on ATP hydrolysis (Zamore et al., 2000), which was subsequently attributed to an energy requirement during the RISC-loading step of the pathway (Kawamata et al., 2009). This led to the "rubber band" model for RISC assembly, where the incorporation of bulky, rigid small-RNA duplexes into Argonaute is facilitated by the ATP-dependent conformational opening of Argonaute (Kawamata and Tomari, 2010). The resulting structural tension is subsequently relieved by the ATP-independent removal of the passenger strand to generate RISC. The ATP-dependent opening of Argonaute is thought to be catalyzed by the Hsc70/Hsp90 machinery, which associate with Argonaute proteins and whose inhibition blocks RISC assembly (Tahbaz et al., 2001; Hock et al., 2007; Landthaler et al., 2008; Iki et al., 2010; Iwasaki et al., 2010).

Following RISC loading, the passenger strand is nicked by Argonaute and the cleavage fragments are removed, a process known as RISC activation (Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). In *Neurospora*, RISC activation is facilitated by the QIP[5] exonuclease,

---

[5] An acryonym for "QDE-2-Interacting Protein", in reference to its physical assocation with the *Neurospora* Argonaute protein QDE-2.

which degrades the passenger-strand fragments (Maiti et al., 2007). The C3PO endonuclease, a complex of Trax and Translin proteins, is thought to perform an analogous function in the *Drosophila* and human RNAi pathways (Liu et al., 2009; Tian et al., 2011; Ye et al., 2011) (Figure 1). Based on the structural similarity between an Argonaute-bound siRNA duplex and Argonaute-bound guide–target duplex, RISC activation by passenger-strand cleavage is thought to be mechanistically similar to target-RNA cleavage (Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). However, a separate set of cofactors have been identified that catalyze removal or decay of target-RNA cleavage fragments, including the La autoantigen, the 5′-3′ exonuclease Xrn1, and the exosome (Orban and Izaurralde, 2005; Liu et al., 2011).

Prior to or following incorporation into RISC, the small RNAs can be modified by RNAi cofactors (Figure 1). *Drosophila* siRNAs contain a 2′-*O*-methyl group on their 3′ termini, a modification that is catalyzed by the *S*-adenosylmethionine–dependent methyltransferase Hen1 acting on the single-stranded guide RNA in RISC (Horwich et al., 2007). *Arabidopsis* siRNAs are also 2′-*O*-methylated, but the substrate for *Arabidopsis* HEN1 is the siRNA duplex (Yang et al., 2006). In both cases, methylation protects the small RNA from 3′-terminal modifications (i.e., trimming or tailing) that can affect the stability of the small RNA (Li et al., 2005; Ameres et al., 2010). Small RNAs can also recruit RNA-dependent RNA polymerase (RdRP) enzymes, which amplify the silencing signal in fission yeast[6], nematodes, and plants by generating additional dsRNA (Smardon et al., 2000; Sijen et al., 2001; Motamedi et al., 2004; Zhang et al., 2007).

RNase III enzymes

The identification of Dicer as the enzyme responsible for siRNA generation was facilitated by extensive prior knowledge about the RNase III family of enzymes, a collection of

---

[6] In fission yeast, the RdRP is absolutely required for the generation of dsRNA-derived siRNAs.

dsRNA-specific endoribonucleases (MacRae and Doudna, 2007). Members of the family contain at least one characteristic ribonuclease domain, referred to as the RNase III domain, which catalyzes the cleavage of dsRNA to yield 2-nt 3′ overhangs with 5′ phosphate and 3′ hydroxyl termini (Robertson et al., 1968). The family is typically subdivided into three classes based on domain composition, i.e., the number of RNase III domains and the presence of additional domains (Lamontagne et al., 2001) (Figure 2A). Class I proteins have a single ribonuclease domain and a dsRBD. In addition to these two domains, class II enzymes have a second RNase III domain and an extended amino-terminal region that lacks any recognizable domains[7]. Class III proteins, which also contain a pair of RNase III domains and a dsRBD, are distinguished from class II enzymes by the presence of a PAZ domain and also typically contain an N-terminal helicase domain followed by a small domain of unknown function (DUF283).

Class I enzymes are found in bacteria, bacteriophage, and fungi where they function in the maturation of ribosomal RNA (rRNA) and other noncoding RNAs (Lamontagne et al., 2001). These RNase III enzymes function as homodimers, with the dimerization interface formed by the RNase III domains and each RNase III domain cleaving one strand of the substrate (Dunn, 1976). Typical substrates for bacterial RNase III are hairpin structures formed by the folding back of two complementary segments, which in the case of the pre-rRNA substrate are separated by ~1700 nucleotides (Young and Steitz, 1978). Although bacterial RNase III enzymes can cleave any sufficiently long dsRNA substrates *in vitro*, there are only a limited number of *in vivo* substrates for these enzymes and cleavage occurs in a site-specific manner (Robertson, 1982; Afonyushkin et al., 2005; Deltcheva et al., 2011; Viegas et al., 2011). The principles of substrate recognition are still not completely understood, but some anti-determinant sequences that disfavor cleavage have been identified (Zhang and Nicholson, 1997). Bacterial RNase III is

---

[7] But contains a proline-rich region (Figure 2A)

thought to recognize the structure of the sugar–phosphate backbone, in part through binding of the dsRBDs, which in the post-catalytic complex sit on the opposite side of the RNA duplex as the RNase III domains (Gan et al., 2006; Pertzev and Nicholson, 2006).

Unlike bacterial RNase III, the *S. cerevisiae* homolog Rnt1 has a well-defined substrate preference: dsRNA hairpins that are capped by a tetraloop with the consensus sequence 'AGNN' (Chanfreau et al., 2000; Nagel and Ares, 2000). Substrate specificity is mediated by the dsRBD, which makes extensive contacts with the RNA and recognizes the distinct geometry of the tetraloop minor groove rather than the sequence of the tetraloop itself (Wu et al., 2001; Leulliot et al., 2004; Wu et al., 2004). Tetraloop binding by the dsRBD positions the pair of active sites in the stem region 13–16 base pairs away for precise cleavage. The major substrates for *S. cerevisiae* Rnt1 are pre-rRNAs[8], small nuclear RNAs, and small nucleolar RNAs (Elela et al., 1996; Chanfreau et al., 1998a; Chanfreau et al., 1998b; Kufel et al., 1999; Qu et al., 1999). The fission yeast *S. pombe* also contains a class I enzyme, Pac1, that processes pre-rRNA, but substrate recognition does not depend on an 'AGNN' tetraloop (Rotondo et al., 1997).

Class II enzymes include the nuclear protein Drosha, which is required for microRNA biogenesis in animals and has also been implicated in pre-rRNA processing in humans (Wu et al., 2000; Lee et al., 2003). Drosha functions as a monomer, with its pair of RNase III domains forming an intramolecular dimer to enact dsRNA cleavage (Han et al., 2004). Unlike class I enzymes, Drosha requires a protein cofactor to accurately cleave its substrates. This specificity is imparted by the dsRNA-binding protein DGCR8[9], which together with Drosha forms the

---

[8] Although pre-rRNA is a substrate for both *E. coli* RNase III and *S. cerevisiae* Rnt1, the processing sites are quite different for the bacterial and yeast substrates. The base-paired region in *E. coli* pre-rRNA is formed by long-range interactions, with the mature rRNA contained within a large loop. The hairpin in *S. cerevisiae* pre-rRNA is formed by locally restricted interactions at the 3′ end of the transcript, with all mature rRNAs located upstream of the cleavage site (Elela et al., 1996).

[9] Pasha (short for "Partner of Drosha") is the *C. elegans* ortholog.

Microprocessor complex (Denli et al., 2004; Gregory et al., 2004). In the microRNA pathway, which is described in greater detail below, the Microprocessor precisely excises pre-microRNA hairpins from longer primary transcripts (Kim et al., 2009). DGCR8 recognizes the junction between the dsRNA stem and single-stranded RNA (ssRNA) flanking regions at the base of the hairpin, which positions the Drosha active sites one helical turn away for accurate cleavage (Han et al., 2006).

Mechanism of Dicer processing

Class III enzymes include the canonical Dicers[10] that are found in plants, animals, and most fungi (MacRae and Doudna, 2007). The discovery of Dicer as the RNase III enzyme that mediates the initiator step of RNAi was followed closely by the first crystal structure of an RNase III enzyme (Blaszczyk et al., 2001), which served as a basis for mechanistic models of Dicer function. Although bacterial RNase III had been isolated and characterized in 1968 (Robertson et al., 1968), the mechanism by which it generated staggered cuts on dsRNA had remained a mystery. Consistent with early studies (Dunn, 1976), the crystal structure of the catalytic domain of *Aquifex aeolicus* RNase III revealed a compact homodimer (Blaszczyk et al., 2001). Based on structural and genetic data, a model for dsRNA cleavage was proposed in which a dsRNA-binding cleft is flanked by two compound catalytic centers, with each cleaving both strands of the dsRNA (Figure 2B). The resulting four cuts would generate 9-nt products with 2-nt 3′ overhangs, which is consistent with the products observed after complete digestion with RNase III *in vitro*. In this model, the compound catalytic centers are comprised of two clusters of acidic residues that are contributed by both monomers, with each cluster responsible for the

---

[10] Defined as Dicer enzymes that contain a PAZ domain

cleavage of a single phosphodiester bond on one strand. Notably, only a single divalent metal ion was observed in each catalytic center in the crystal structure.

Dicer differs from bacterial RNase III enzymes in two major respects: Each polypeptide chain of Dicer contains two RNase III domains (designated RNase IIIa and IIIb), and Dicer generates longer products. Based on the proposed model for bacterial RNase III cleavage, models for Dicer cleavage that involved the compound catalytic centers were envisioned that accounted for these differences. Initial models invoked the formation of an intermolecular dimer containing four RNase III domains arranged either in a head-to-tail or antiparallel configuration (Zamore, 2001; Hannon, 2002; Carmell and Hannon, 2004) (Figure 2C). To account for the longer products of Dicer cleavage, it was noted that the RNase IIIb domain contains an amino-acid substitution at a putative catalytic residue that would render the associated active site non-functional. As a result, the Dicer homodimer would contain two functional active sites spaced twice as far apart as those of a bacterial RNase III homodimer, which would explain the generation of ~22-nt products (Figure 2C). An alternative model proposed the formation of an intramolecular pseudodimer between the RNase IIIa and RNase IIIb domains of a single Dicer molecule, with product length determined by the distance between the existing end of the dsRNA substrate and the single active catalytic center (Carmell and Hannon, 2004) (Figure 2D). This model was more consistent with the observation that recombinant human Dicer preferentially cleaves dsRNA substrates at their termini (Zhang et al., 2002).

Efforts to understand the mechanism of Dicer cleavage eventually called into question the compound catalytic center. Biochemical analysis of Dicer mutants containing substitutions at residues implicated in catalysis revealed that the RNase IIIa and IIIb active sites can function independently of each other, which was inconsistent with the compound catalytic center (Zhang et al., 2004). These analyses led to a model in which a single processing center was responsible for cleavage, with the active-site residues of RNase IIIa catalyzing the cleavage of one strand and the active-site residues of RNase IIIb catalyzing the cleavage of the opposite strand. Analogous experiments using *E. coli* RNase III also identified a single processing center present within the homodimeric enzyme. Further insight into the molecular mechanism of Dicer cleavage came from the observation that the purified enzyme sedimented as a ~180-kilodalton molecule, which was inconsistent with the formation of an intermolecular dimer (Zhang et al., 2004).

Having established which RNase III domains assemble together to cleave each strand, it remained to be determined how the Dicer active sites are precisely positioned at the cleavage site to generate discretely sized products. Since Dicer preferentially cleaves dsRNA substrates at their termini (Zhang et al., 2002) and most efficiently processes substrates bearing 2-nt 3′ overhangs (Zhang et al., 2004), it appeared likely that Dicer itself recognized the terminus of the dsRNA substrate. The terminus-binding function within Dicer was hypothesized to reside within the PAZ domain on the basis of RNA-binding studies of PAZ domains found in Argonaute-family proteins (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). The impaired processing of dsRNA substrates by Dicer containing mutations in the PAZ domain was consistent a role for this domain in recognizing the substrate terminus (Zhang et al., 2004). The single processing center model for catalysis and PAZ-mediated recognition of the dsRNA terminus together suggested a model for Dicer measurement: Small-RNA products are successively removed from the end of the dsRNA substrate, with product length determined by the distance spanning the terminus-binding PAZ domain and RNase III active sites (Zhang et al., 2004) (Figures 2D and 2E).

The confirmation that Dicer functions as a molecular ruler came with the crystal structure of a simple Dicer enzyme from *Giardia intestinalis* (Macrae et al., 2006). The structure of *Giardia* Dicer revealed an intramolecular RNase III domain dimer with a pair of active sites separated from the PAZ domain by ~65 Å, a distance that matches the ~25-nt length of *Giardia* Dicer products. Between the RNase III and PAZ domains is a flat positively charged surface on which the dsRNA would be positioned for cleavage. Further analysis of *Giardia* Dicer confirmed that the PAZ domain was required to specify product length (MacRae et al., 2007). Moreover, by comparing the processing of substrates bearing different termini geometries (2-nt 3′ overhang, blunt, or 2-nt 5′ overhang), it was shown that *Giardia* Dicer determines product length by measuring from the free 3′ end.

The universality of measurement from the 3′ end by Dicer was recently challenged by the finding that human Dicer can also anchor the 5′ end of a substrate (Park et al., 2011). This 5′ measuring mechanism requires a 5′ monophosphate that is recognized by a novel basic motif within Dicer, the mutation of which impairs substrate processing and alters cleavage sites *in vitro*. Consistently, however, substrates bearing 2-nt 3′ overhangs are processed most efficiently (Park et al., 2011), indicating that the geometry of the overhang itself is a critical determinant of Dicer processing. Nevertheless, the 5′-measuring mechanism is conserved in metazoan Dicers, where it may enable processing of pre-microRNA substrates irrespective of 3′-end modifications that might otherwise shift the Dicer cleavage site (Park et al., 2011).

In addition to the PAZ domain and tandem RNase III domains found in *Giardia* Dicer, most canonical Dicer enzymes also have an N-terminal helicase domain and a pair of double-stranded RNA-binding domains (dsRBDs) (MacRae and Doudna, 2007). The helicase domain, the removal of which stimulates Dicer activity (Ma et al., 2008), contributes to processivity by using ATP hydrolysis to facilitate translocation on long dsRNA substrates (Cenik et al., 2011; Welker et al., 2011). According to this model, processivity results from alternating cycles of cleavage and ATP-dependent translocation without intervening dissociation of the Dicer–dsRNA complex. The most recent electron-microscopy reconstruction of human Dicer[11] localized the helicase domain to a position adjacent to the RNase III active site where it could remain bound to the dsRNA substrate after cleavage, thereby providing the structural basis for this processivity (Lau et al., 2012). Both dsRBDs[12] have dsRNA-binding activity and may therefore contribute to substrate binding, but they have also been shown to be important for subcellular localization and protein–protein interactions in fission-yeast and plant Dicers, respectively (Qin et al., 2010; Barraud et al., 2011).

Molecular architecture of RISC

Argonaute proteins have a conserved domain structure consisting of N, PAZ, MID, and PIWI domains (Song et al., 2004) (Figure 3A). The N domain contributes to small-RNA duplex unwinding during RISC activation, perhaps by functioning as a wedge to actively pry apart the two strands of the duplex (Kwak and Tomari, 2012). The PAZ domain, which is only found in Dicer and Argonaute proteins, is an RNA-binding module that adopts a variant OB[13] fold (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). A highly conserved pocket in the PAZ domain recognizes 2-nt 3′ overhangs in a sequence-independent manner by simultaneously binding the backbone of the overhang-containing strand and capping the 5′ end of the opposite strand (Lingel et al., 2004; Ma et al., 2004). The MID domain adopts a Rossmann-like fold, which is a

---

[11] Determining the crystal structure of a metazoan Dicer remains a challenge in the field.

[12] One of which is technically the DUF283, which adopts a dsRBD fold (Dlakic 2006; Qin et al., 2010)

[13] Oligonucleotide/oligosaccharide-binding fold, in reference to the ligands of the four proteins that were first shown to adopt this common fold (Murzin 1993)

nucleotide-binding structural motif (Song et al., 2004). In RISC, the 5′-terminal nucleotide of the guide RNA is flipped out and recognized by the MID domain (Ma et al., 2005; Parker et al., 2005), consistent with observations that the first position of the guide is not important for targeting (Lewis et al., 2005). Interactions of the first base with a nucleotide specificity loop within the MID domain enable sensing of the 5′-nucleotide identity, which provides a mechanistic basis for the 5′-nucleotide biases of Argonaute-bound guide RNAs[14] and may contribute to small-RNA sorting between Argonaute proteins in some organisms (Boland et al., 2010; Frank et al., 2010; Frank et al., 2012; Zha et al., 2012). The PIWI domain harbors the active site, which is described in more detail below. At the MID–PIWI interface, a highly conserved basic pocket anchors the 5′ phosphate of the guide RNA (Parker et al., 2004; Ma et al., 2005; Parker et al., 2005), consistent with early biochemical studies establishing the importance of this phosphate for RISC assembly *in vitro* (Nykanen et al., 2001).

In RISC, the guide strand lies in a nucleic-acid-binding channel that is positioned between the N–PAZ and MID–PIWI lobes (Wang et al., 2008b). With the exception of the 5′ nucleotide, the bases at the 5′ end of the guide are stacked in a helical conformation with Watson–Crick faces exposed to solvent. This pre-organization[15] of the 5′ end of the guide provides a structural basis for the disproportionate contribution of this “seed region” of the guide RNA to target binding (Lewis et al., 2003; Bartel, 2004; Doench and Sharp, 2004; Haley and Zamore, 2004). Target binding leads to a slight widening of the nucleic-acid-binding channel and the formation of an A-form duplex between the seed region and the target RNA (Wang et al., 2008a). In both the binary complex and ternary complexes with short target RNAs, the 3′ end of

---

[14] Argonaute-family proteins display a preference for certain nucleotides at the 5′-ends of guide RNAs, which can influence guide-strand selection. This 5′-nucleotide bias differs among Argonaute-family proteins.

[15] Although *pre-organized* with respect to the guide–target duplex, the helical geometry of the 5′ end of the guide RNA is *maintained* with respect to the siRNA duplex from which it was derived.

the guide strand is bound to the PAZ domain while the 5′ end of the guide strand is anchored in the 5′-nucleotide binding pocket (Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009b). However, the 3′ end of the guide strand is released from the PAZ domain when longer target RNA is present and guide–target pairing is propagated beyond one helical turn (Wang et al., 2009b).

These structural observations are consistent with a ‘two-state model’ for target recognition (Bartel, 2004; Tomari and Zamore, 2005). During the nucleation step, the guide strand is anchored at both ends and annealing with the target RNA is restricted to the seed region due to topological constraints. During the propagation step, the 3′ end of the guide strand is released from the PAZ domain and base pairing is allowed to extend to the 3′ end of the guide. The transition from the nucleation to propagation steps is accompanied by pivot-like domain movements within Argonaute (Wang et al., 2009b). In the resulting guide–target duplex, the phosphate of the target RNA across from nucleotides 10–11of the guide strand is positioned at the active site for cleavage.

Argonaute active site

RNase H utilizes a two-metal-ion mechanism to catalyze phosphodiester bond hydrolysis (Steitz and Steitz, 1993), with the divalent metal ions stabilized in part by a collection of carboxylate residues at the active site (Nowotny and Yang, 2006). Structural homology between the PIWI domain and RNase H identified a pair of highly conserved aspartate residues in Argonaute that are located at the same positions as the invariant carboxylates in the active site of the RNase H fold (Parker et al., 2004; Song et al., 2004). A third conserved carboxylate that was in close proximity was initially proposed to constitute the last component of a ‘DDE’ catalytic triad that mediated Argonaute cleavage (Liu et al., 2004; Song et al., 2004), analogous to the 'DDE' active site of other members of the RNase H superfamily (Yang and Steitz, 1995). However, mutation of this third carboxylate in Argonaute was subsequently found to not affect cleavage activity; instead a nearby histidine residue was identified as the missing active-site residue of a 'DDH' catalytic triad[16] (Rivas et al., 2005).

Despite its broad acceptance, the Argonaute catalytic triad is not sufficient to explain the metal-ion coordination seen in the active site of RNase H, which is mediated by a 'DEDD' catalytic tetrad (Hall, 2005; Nowotny et al., 2005) (Figure 3B). Other conserved residues near the Argonaute active site have been proposed as possible substitutes for the catalytic glutamate, including a distantly positioned glutamate and a nearby arginine (Song et al., 2004; Hall, 2005; Nowotny et al., 2005; Yuan et al., 2005), but functional analysis of the glutamate residue has been inconsistent with a role in catalysis (Rivas et al., 2005). Nevertheless, based on the conservation of active-site geometry and catalytic mechanism between RNase H and Argonaute, it has been suggested that a fourth residue may contribute to catalysis but may only be properly positioned in the active form of the enzyme (Hall, 2005; Nowotny et al., 2005).

Part 3: Diversity of RNA-silencing pathways

Although RNAi was discovered as a gene-silencing pathway triggered by exogenous dsRNA, it was immediately recognized that the RNAi machinery was likely to mediate physiological forms of gene silencing (Fire et al., 1998). Indeed, RNAi turned out to be just one member of a collection of gene-silencing pathways that are mediated by a small-RNA guide bound to a member of the Argonaute superfamily (Hutvagner and Simard, 2008). RNAi is

---

[16] 'DDD' catalytic triads are also common in Argonaute proteins with cleavage activity, e.g. bacterial Argonautes.

believed to be the evolutionary ancestor of all of these RNA-silencing pathways (Shabalina and Koonin, 2008).

microRNAs

MicroRNAs (miRNAs) are a class of ~22-nt RNAs that are distinguished from siRNAs based on their processing from ssRNA precursors that fold into characteristic stem-loop structures (Bartel, 2004). The first miRNA *lin-4* was identified in *C. elegans* in the early 1990s (Lee et al., 1993; Wightman et al., 1993), which was followed seven years later by the discovery of the second miRNA[17] *let-7* (Reinhart et al., 2000). Unlike *lin-4*, which is only found in worms, *let-7* was found to be conserved among bilaterians (Pasquinelli et al., 2000), which provided the first indication that miRNAs were not simply a worm-specific oddity. Still, that *lin-4* and *let-7* represented only the founding members of an enormous class of regulatory RNAs was not appreciated until the first small-RNA cloning experiments in worms, flies, and mammals revealed nearly a hundred miRNAs (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Ruvkun, 2001). Since then, hundreds more microRNAs have been identified in animals and plants (Kim and Nam, 2006).

The biogenesis of canonical animal miRNAs begins in the nucleus where a primary miRNA transcript containing a stem-loop structure is processed by the Microprocessor, comprising the RNase III enzyme Drosha and the dsRNA-binding protein DGCR8, to generate a pre-miRNA hairpin (Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004) (Figure 4A, top). Following export to the cytoplasm by Exportin-5 (Lund et al., 2004), the pre-miRNA is cleaved by Dicer[18] to generate a miRNA–miRNA* duplex (Lee

---

[17] At the time, these RNAs were referred to as small temporal RNAs (stRNAs).

[18] In complex with a dsRNA-binding protein (e.g., TRBP in humans)

et al., 2002). The duplex is then asymmetrically incorporated into an Argonaute protein (Khvorova et al., 2003; Schwarz et al., 2003), wherein the duplex is unwound by a mysterious helicase-like activity and the miRNA* strand is discarded (Kawamata and Tomari, 2010). Plant miRNA biogenesis differs from that described above in that both RNase III cleavage events are mediated by a single Dicer-like protein (DCL1) in the nucleus (Park et al., 2002; Reinhart et al., 2002; Papp et al., 2003; Kurihara and Watanabe, 2004) (Figure 4A, bottom).

Animal miRNAs primarily bind mRNA targets in the 3′ untranslated region based on a short region of complementarity to the 5′ end of the miRNA (which corresponds to the previously discussed "seed region") (Lewis et al., 2003). Although instances of extensive complementarity can lead to target cleavage by some Argonautes[19] (Yekta et al., 2004), repression by animal miRNAs is generally mediated by mRNA destabilization or translational repression[20] (Guo et al., 2010; Bazzini et al., 2012; Djuranovic et al., 2012). In contrast, plant miRNAs have extensive complementarity to their targets and typically guide cleavage by Argonaute (Llave et al., 2002; Rhoades et al., 2002; Tang et al., 2003). These notable differences between the plant and animal miRNA pathways have suggested independent convergent evolution of the pathways, which is further supported by the absence of homologous miRNAs between plants and animals (Jones-Rhoades et al., 2006; Axtell et al., 2011).

piRNAs

Piwi-interacting RNAs (piRNAs) are unique among small-RNA guides in being generated from ssRNA precursors in a Dicer- and RdRP-independent manner (Vagin et al.,

---

[19] Only some animal Argonautes have cleavage activity.

[20] The relative contributions of these mechanisms remains a subject of debate.

2006). piRNAs were first identified in mammalian testes based on their association with Piwi[21] proteins (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2006), a subdivision of the Argonaute family of proteins with restricted expression primarily in the germline and stem cells (Aravin et al., 2007). An analogous class of Piwi-associated RNAs was subsequently identified in *C. elegans*, *Drosophila*[22], and zebrafish (Ruby et al., 2006; Vagin et al., 2006; Brennecke et al., 2007; Houwing et al., 2007; Yin and Lin, 2007; Batista et al., 2008; Das et al., 2008).

The majority of piRNAs are derived from discrete genomic loci that correspond to repetitive elements (Malone and Hannon, 2009). In contrast to the Dicer-dependent generation of most other classes of small RNAs, piRNAs[23] are generated by a ping-pong mechanism in which antisense piRNAs guide the Piwi-dependent cleavage of sense transcripts (and vice versa), with each cleavage event generating the 5′ end of a new piRNA (Brennecke et al., 2007; Gunawardane et al., 2007) (Figure 4B). The piRNA pathway constitutes an adaptive immune system that protects the genome against transposable elements by silencing them (Siomi et al., 2011).

Endogenous siRNAs

Just as exogenous dsRNA can be processed into siRNAs that mediate RNAi, endogenous sources of dsRNA can similarly give rise to active siRNA species, known as endogenous siRNAs (endo-siRNAs). The dsRNA substrate can be generated by an RdRP acting on an ssRNA

---

[21] An abbreviation for "P-element induced wimpy testis" in reference to the phenotype of *Drosophila* mutants (Lin and Spradling, 1997)

[22] Originally identified as repeat associated siRNAs, or rasiRNAs (Aravin et al., 2001)

[23] The ping-pong mechanism applies to the biogenesis of secondary piRNAs. The mechanism by which primary piRNAs are generated is still largely unknown but likely involves a long ssRNA precursor that is endonucleolytically cleaved to generate the 5′ end of the mature piRNA and exonucleolytically trimmed back to generate the 3′ end.

template, as is the case for trans-acting siRNAs in plants, heterochromatic siRNAs in fungi (described below), and secondary siRNAs[24] in nematodes (Reinhart and Bartel, 2002; Peragine et al., 2004; Vazquez et al., 2004b; Sijen et al., 2007). For many years, it was thought that endo-siRNAs were restricted to fungi, plants, and nematodes, each of which encodes an RdRP. This view was subsequently overturned with the discovery of endo-siRNAs in *Drosophila* and mouse oocytes (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a; Okamura et al., 2008b; Tam et al., 2008; Watanabe et al., 2008). These endo-siRNAs can be derived from the pairing of endogenous sense and antisense transcripts, as is also the case for natural siRNAs in plants (Borsani et al., 2005). Alternative sources of duplex substrates include ssRNA transcripts that form long hairpin structures, as for *Drosophila* hairpin RNAs, and convergent transcripts that anneal at their 3′ ends (Ghildiyal and Zamore, 2009).

***Heterochromatic siRNAs in* S. pombe**

The fission yeast *Schizosaccharomyces pombe* contains a single homolog each of Dicer, Argonaute, and RdRP (Cerutti et al., 2000; Provost et al., 2002; Wood et al., 2002). Cloning of endogenous small RNAs from *S. pombe* identified a population of 20–25-nt RNAs mapping to the heterochromatic centromeric repeats (Reinhart and Bartel, 2002). In parallel, analysis of strains lacking each of the RNAi components revealed loss of heterochromatin at the centromeres and concomitant accumulation of centromeric transcripts (Volpe et al., 2002). Together, these studies implicated the RNAi machinery in heterochromatin assembly and silencing at centromeric repeats in *S. pombe*. Heterochromatin formation at the silent mating-type locus of fission yeast also depends on the RNAi machinery (Hall et al., 2002).

---

[24] In contrast to the primary siRNAs derived from the trigger dsRNA

S. pombe Argonaute (Ago1) interfaces with the chromatin machinery through the RNA-induced transcriptional silencing (RITS) complex, comprised of centromeric siRNAs, Ago1, Chp1, and Tas3 (Verdel et al., 2004). Chp1 is a chromodomain-containing protein that interacts with methylated lysine 9 on histone H3 (Partridge et al., 2002; Schalch et al., 2009), while the GW-motif-containing protein Tas3 provides a physical link between Ago1 and Chp1 (Debeauchamp et al., 2008). Ago1 forms a separate complex with Arb1 and Arb2 termed the Argonaute siRNA chaperone (ARC) complex, which contains primarily duplex siRNAs and may function in transferring siRNA products from Dcr1 into Ago1[25] (Buker et al., 2007). The RNA-dependent RNA polymerase Rdp1 also exists in a complex, in this case with the RNA helicase Hrr1 and the poly(A) polymerase Cid12 (Motamedi et al., 2004). This so-called RDRC (RNA-dependent RNA polymerase complex) has RdRP activity and is localized in the nucleus with and interacts with RITS.

Both RITS and RDRC are associated with centromeric noncoding transcripts, suggesting that the Ago–siRNA complex base-pairs with nascent transcripts rather than with the DNA itself (Motamedi et al., 2004). Consistent with an important role for nascent transcripts, RNAi-dependent heterochromatin assembly is affected by mutations in subunits of RNA polymerase II (Djupedal et al., 2005; Kato et al., 2005), and spreading of heterochromatin requires read-through transcription (Irvine et al., 2006). Furthermore, artificial tethering of the RITS complex to an RNA transcript is sufficient to initiate RNAi- and heterochromatin-dependent silencing (Buhler et al., 2006).

Once bound to a nascent homologous transcript, RITS associates with Stc1, which recruits the chromatin-modifying CLRC complex (Bayne et al., 2010). Methylation of H3K9 by Clr4, a component of the CLRC complex, creates a binding site for the Chp1 component of RITS

---

[25] ARC may be functionally analogous to the RISC-loading complex.

and thereby stabilizes the association of RITS with chromatin (Partridge et al., 2002). RITS also recruits RDRC to generate dsRNA that is processed into siRNAs by Dcr1, thus amplifying the process (Sugiyama et al., 2005). The initial recruitment of RITS is mediated by a population of Dcr1- and Rdp1-independent small RNAs derived from abundant transcripts, which are known as primal RNAs (Halic and Moazed, 2010). The ultimate outcome of this complex cascade of events is the formation of heterochromatin at RITS-associated loci.

Part 4: RNAi in budding yeast?

The first indication that the model budding yeast *Saccharomyces cerevisiae* lacks an RNAi pathway predates[26] the discovery of RNAi itself. When the *AGO1* locus was cloned from *Arabidopsis thaliana*, it was found to encode a 115-kilodalton protein that had high sequence similarity to a large number of putative proteins in many plants and animals (Bohmert et al., 1998). Notably, however, the AGO1 protein had no similarity to any publicly available yeast sequences[27], which at the time were primarily from *S. cerevisiae* owing to its recently sequenced genome (Goffeau et al., 1996). Following the discovery of RNAi and the role of RdRP in the *S. pombe* pathway, the notable absence of Argonaute from *S. cerevisiae* was extended to include the absence of RdRP and interpreted as the absence of an RNAi pathway (Aravind et al., 2000; Cerutti et al., 2000). The identification of Dicer as a component of the pathway similarly led to a realization that *S. cerevisiae* lacks an identifiable homolog (Provost et al., 2002).

With *S. cerevisiae* reigning as the model budding-yeast species, the fact that RNAi is missing in *S. cerevisiae* became generalized to the statement that RNAi is absent from budding

---

[26] By only six weeks

[27] The absence of AGO1 homologs in bacterial sequences was also noted at the time. As with yeast genome sequences, the subsequent availability of many more bacterial genome sequences revealed that some bacteria do, in fact, encode proteins of the Argonaute family.

yeasts (e.g., Stoica et al., 2006). The first hint that this generalization may not be accurate came from sequencing the genomes of *Saccharomyces castellii* and *Candida albicans*, which revealed potential Argonaute homologs in each case (Cliften et al., 2003; Jones et al., 2004; Axelson-Fisk and Sunnerhagen, 2006; Nakayashiki et al., 2006). The same Argonaute-only repertoire of RNAi proteins was later found in the genome sequence of *Kluyveromyces polysporus*[28], a close relative of *S. castellii* (Scannell et al., 2007). However, the absence of obvious Dicer and RdRP homologs in budding-yeast genomes led to the suggestion that these Argonaute proteins may perform non-RNAi functions (Axelson-Fisk and Sunnerhagen, 2006). Indeed, the absence of Dicer was particularly puzzling, as RdRP-independent RNAi pathways are prevalent in eukaryotes but no Dicer-independent RNAi pathways[29] are known (Obbard et al., 2009). Certain prokaryotes also have Argonaute homologs yet lack other genes associated with RNA-silencing pathways (Hall, 2005), suggesting that the budding-yeast and bacterial Argonautes may perform similar (yet undefined) non-RNAi functions. Still, the Argonaute-encoding budding yeasts were mostly ignored as isolated examples, with many more budding-yeast genomes resembling that of *S. cerevisiae* with respect to the absence of all RNAi machinery (Axelson-Fisk and Sunnerhagen, 2006).

These observations led us to wonder about the role of Argonaute in these exceptional budding-yeast species. In particular, are they really missing an RNA-silencing pathway? It is this question with which my thesis research began.

---

[28] The species name refers to the fact that its asci typically contain 50–100 spores as a result of post-meiotic mitotic replications (van der Walt, 1956).

[29] In the strictest sense as defined by Fire et al., 1998

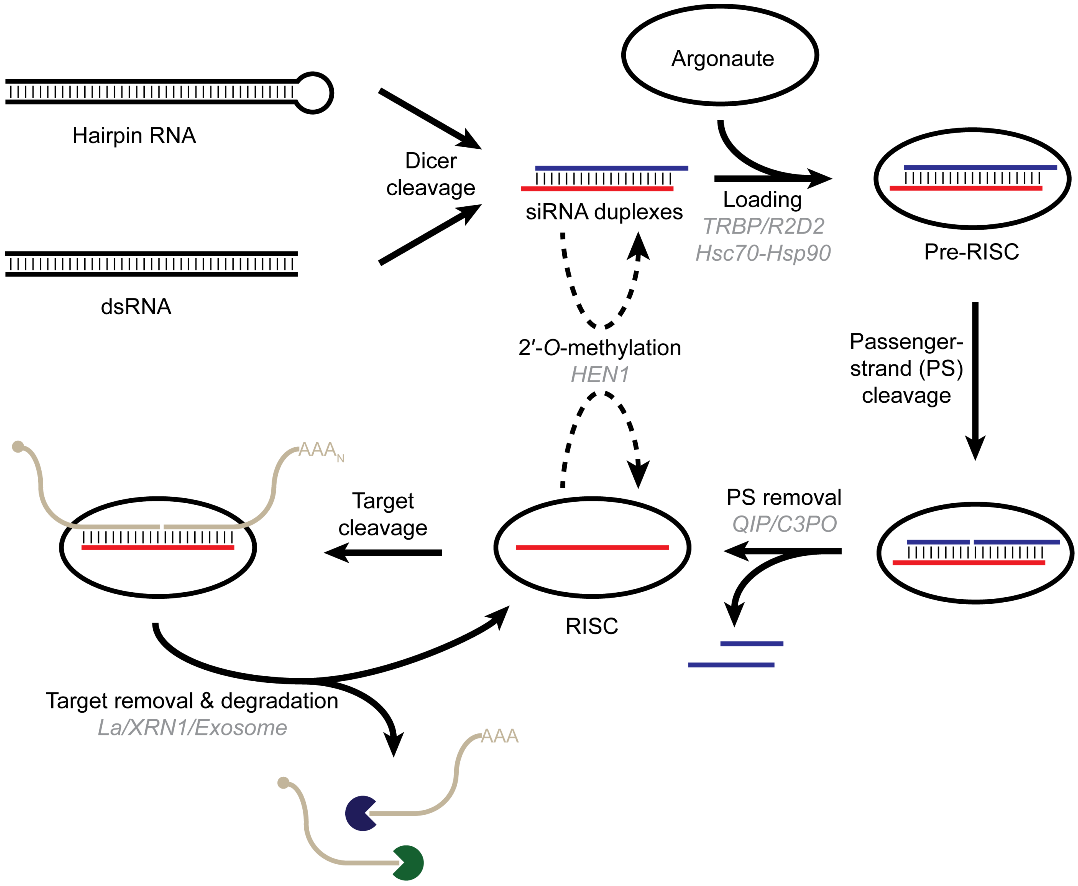


Figure 1. Overview of the RNAi pathway
RNAi is initiated when a double-stranded RNA is cleaved by Dicer to generate siRNA duplexes bearing 2-nt 3′ overhangs. These duplexes are loaded into Argonaute to form the pre-RISC complex. The passenger strand (shown in blue) is subsequently cleaved by the endonucleolytic activity of Argonaute and discarded. The resulting mature RISC binds and cleave targets that have extensive complementarity to the guide RNA (shown in red). Multiple turnover of RISC requires that the resulting cleavage fragments (shown in tan) be removed, after which they are degraded by cellular nucleases. Representative cofactors that have been implicated in the RNAi pathway are shown in gray italics. The 2′-*O*-methylation of siRNAs, depicted as a dashed line, only occurs in certain organisms.

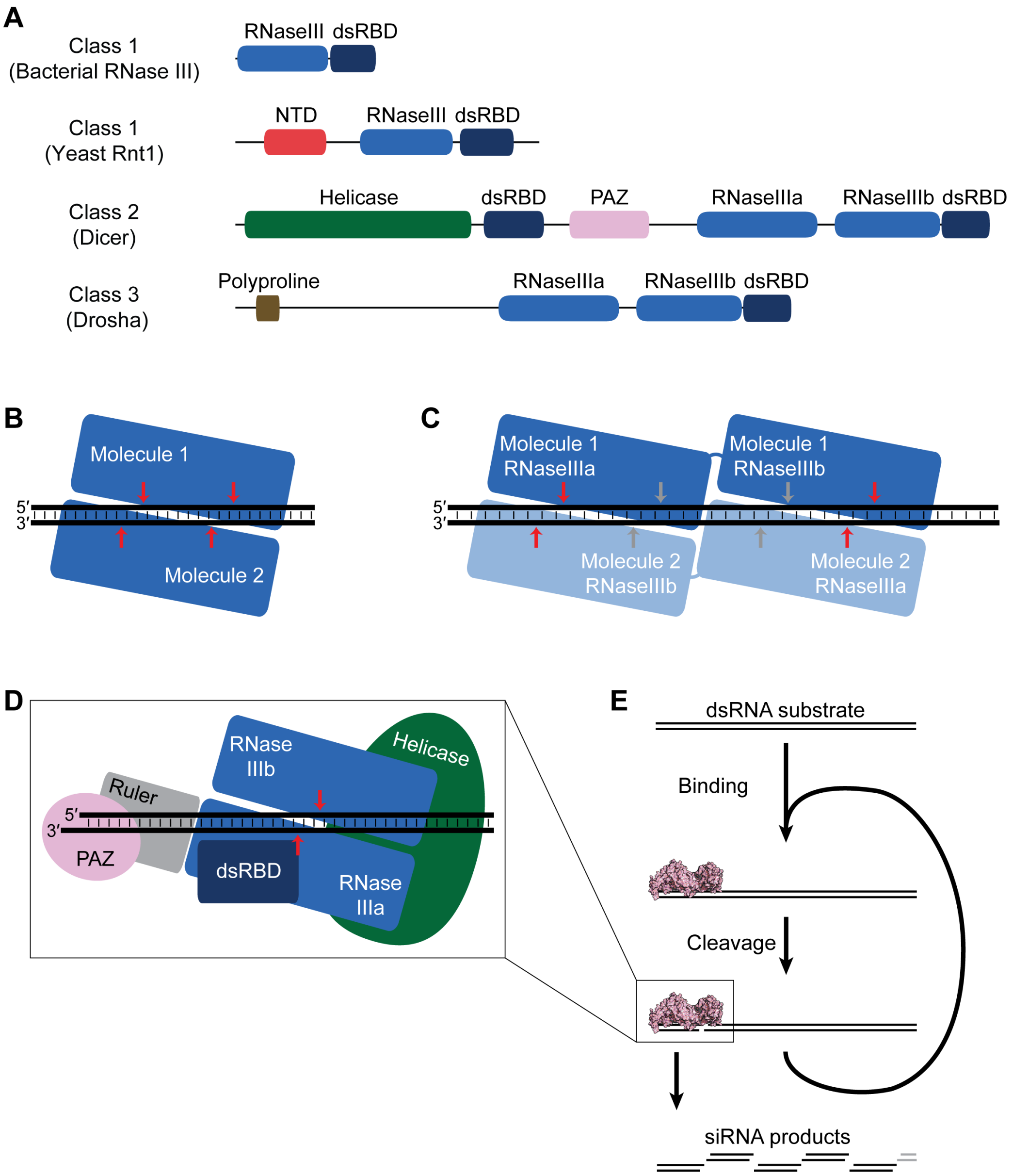


Figure 2. Classification and mechanisms of RNase III enzymes
(A) Classification of RNase III enzymes (adapted from MacRae and Doudna, 2007). Shown are representative domain architectures for each class of RNase III enzymes, with the corresponding proteins indicated in parentheses. NTD, N-terminal domain unique to yeast RNase III enzymes.
(B, C) Models for dsRNA cleavage by bacterial RNase III (B) and canonical Dicer (C). Initial models of bacterial RNase III proposed that each end of the dsRNA-binding cleft contained an active site that cleaved both strands of the dsRNA substrate, thereby resulting in four cuts (red arrows). Based on this model, it was proposed that Dicer formed an intermolecular dimer containing a total of four RNase III domains and four potential active sites. However, the middle pair of active sites were thought to be non-functional (gray arrows) due to an amino-acid substitution at a putative catalytic residue in the RNase IIIb domain.
(D) Measuring product length (based on Zhang et al., 2004 and Lau et al., 2012). Canonical Dicers measure product length from the dsRNA terminus by anchoring the existing 2-nt 3′

overhang in the PAZ domain. This positions the RNase III active sites at a precise distance from the dsRNA terminus, which is determined by the size of the “Ruler” domain. The relative orientations of the dsRBD and helicase domain are modeled after Lau et al., 2012. The cleavage sites, separated by a 2-nt 3′ overhang, are shown as red arrows.

(E) Processive substrate cleavage by canonical Dicers. The structure shown is the minimal Dicer from *Giardia intestinalis* that lacks a helicase domain (Macrae et al., 2006). Processive cleavage across the length of a dsRNA substrate is achieved by cycles of binding and cleavage coupled with movement along the substrate, which is facilitated by the ATP-dependent helicase domain.

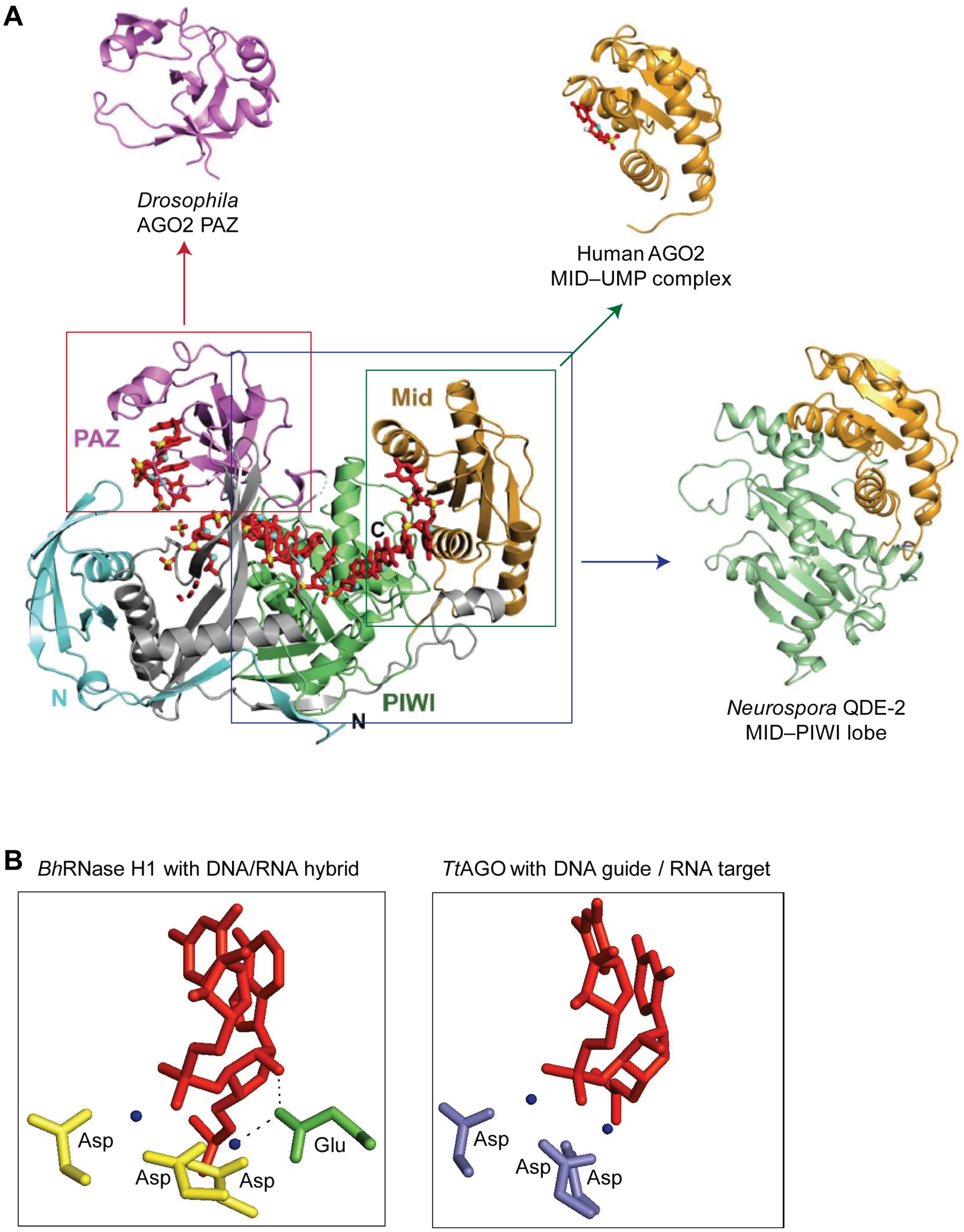


Figure 3. Structural basis for RNA binding and catalysis by Argonaute
(A) Structural analysis of Argonaute proteins. The structure of a bacterial Argonaute with guide DNA (bottom left; reproduced from Wang et al., 2008b) is shown as a ribbon diagram, with domains colored as indicated. Boxed regions correspond to fragments of Argonaute for which a crystal structure of the eukaryotic domain or lobe was known: the PAZ domain of *Drosophila* AGO2 (Lingel et al., 2003), which showed that it adopts an OB fold; the MID domain of human AGO2 bound to uridine monophosphate (UMP, which mimics the 5′ nucleotide of the guide RNA) (Frank et al., 2010), which revealed the structural basis for 5′-nucleotide specificity; and the MID–PIWI lobe of the *Neurospora crassa* Argonaute homolog QDE-2 (Boland et al., 2011),

which represents the largest fragment of eukaryotic Argonaute that had been structurally characterized.

(B) Comparative analysis of RNase H–like active sites. Close-up views around the active sites of *Bacillus halodurans* RNase H1 (left, Nowotny et al., 2005) and *Thermus thermophilus* Argonaute (right, Wang et al., 2009). Nucleotides flanking the scissile phosphate are shown in red, divalent metal ions are depicted as blue spheres, and known active-site residues are drawn in stick representation.

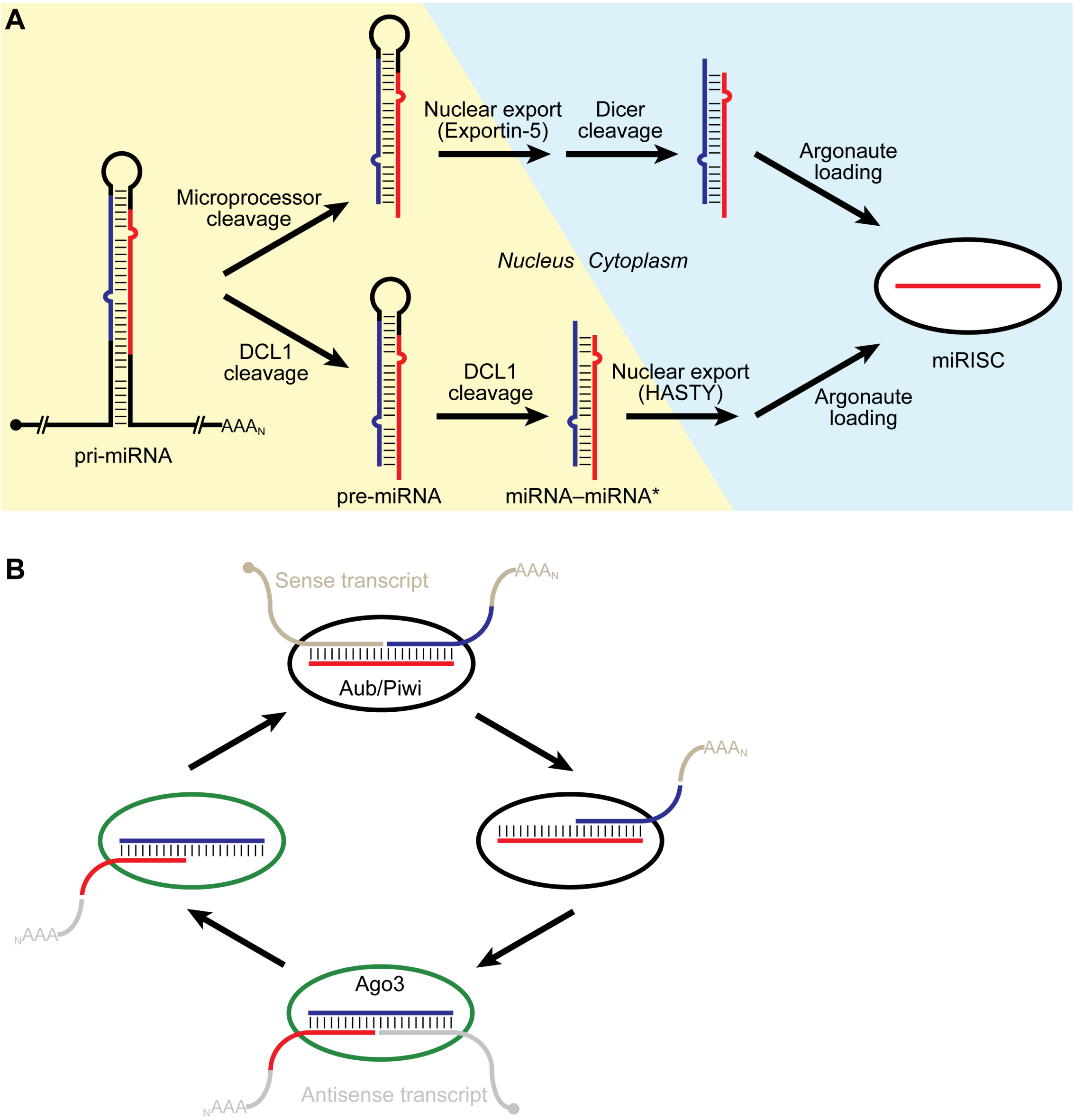


Figure 4. Biogenesis of miRNAs and piRNAs

(A) miRNA biogenesis in animals (top) and plants (bottom). A capped and polyadenylated primary miRNA (pri-miRNA) transcript undergoes two sequential cleavage events to yield a miRNA–miRNA* duplex, which is loaded into Argonaute to yield the mature miRISC. In animals, the nuclear Microprocessor complex generates the pre-miRNA, which is exported to the cytoplasm by Exportin-5 and subsequently cleaved by Dicer to yield the miRNA–miRNA* duplex. In plants, a single enzyme (Dicer-like 1, DCL1) performs both cleavage events in the nucleus prior to export of the miRNA–miRNA* duplex by the Exportin-5-like protein HASTY. (B) Ping-pong mechanism of piRNA biogenesis. An antisense piRNA (red) loaded in Aubergine (Aub) or Piwi guides the cleavage of a sense transcript, which generates the 5′ end of a new sense piRNA (blue). Following 3′-end formation by an unknown mechanism, the sense piRNA is loaded into Ago3 and guides the cleavage of an antisense transcript (gray) to generate the 5′ end of an antisense piRNA. The antisense piRNA undergoes 3′-end formation and is loaded into Aubergine or Piwi, thus restarting the amplification cycle.

References

Afonyushkin, T., Vecerek, B., Moll, I., Blasi, U., and Kaberdin, V.R. (2005). Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB. Nucleic Acids Res *33*, 1678-1689.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761-764.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol *11*, 1017-1027.

Aravind, L., Watanabe, H., Lipman, D.J., and Koonin, E.V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A *97*, 11319-11324.

Axelson-Fisk, M., and Sunnerhagen, P. (2006). Comparative Genomics: Using Fungi as Models. In Topics in Current Genetics (Heidelberg Germany, Springer), pp. 1-28.

Axtell, M.J., Westholm, J.O., and Lai, E.C. (2011). Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol *12*, 221.

Barraud, P., Emmerth, S., Shimada, Y., Hotz, H.R., Allain, F.H., and Buhler, M. (2011). An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. EMBO J *30*, 4223-4235.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bass, B.L. (2000). Double-stranded RNA as a template for gene silencing. Cell *101*, 235-238.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., *et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.L., Kim, D.U., Park, H.O., Ponting, C.P., Rappsilber, J., *et al.* (2010). Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell *140*, 666-677.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-237.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure *9*, 1225-1236.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J *17*, 170-180.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep *11*, 522-527.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell *123*, 1279-1291.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873-886.

Buker, S.M., Iida, T., Buhler, M., Villen, J., Gygi, S.P., Nakayama, J., and Moazed, D. (2007). Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. Nat Struct Mol Biol *14*, 200-207.

Carmell, M.A., and Hannon, G.J. (2004). RNase III enzymes and the initiation of gene silencing. Nat Struct Mol Biol *11*, 214-218.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature *404*, 245.

Caudy, A.A., Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M., Tops, B.B., Silva, J.M., Myers, M.M., Hannon, G.J., and Plasterk, R.H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. Nature *425*, 411-414.

Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. Mol Cell.

Cerutti, L., Mian, N., and Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem Sci *25*, 481-482.

Chanfreau, G., Buckle, M., and Jacquier, A. (2000). Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. Proc Natl Acad Sci U S A *97*, 3142-3147.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998a). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. J Mol Biol *284*, 975-988.

Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998b). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J *17*, 3726-3737.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Cogoni, C., and Macino, G. (1997). Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa. Proc Natl Acad Sci U S A *94*, 10233-10238.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., *et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., *et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell *31*, 79-90.

Debeauchamp, J.L., Moses, A., Noffsinger, V.J., Ulrich, D.L., Job, G., Kosinski, A.M., and Partridge, J.F. (2008). Chp1-Tas3 interaction is required to recruit RITS to fission yeast centromeres and for maintenance of centromeric heterochromatin. Mol Cell Biol *28*, 2154-2166.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev *19*, 2301-2306.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science *336*, 237-240.

Dlakic, M. (2006). DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. Bioinformatics *22*, 2711-2714.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

Dunn, J.J. (1976). RNase III cleavage of single-stranded RNA. Effect of ionic strength on the fidelity of cleavage. J Biol Chem *251*, 3807-3814.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

Elela, S.A., Igel, H., and Ares, M., Jr. (1996). RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. Cell *85*, 115-124.

Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. EMBO J *31*, 3588-3595.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Gan, J., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S., and Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. Cell *124*, 355-366.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., *et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. Genes Dev *20*, 1709-1714.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol *11*, 599-606.

Halic, M., and Moazed, D. (2010). Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell *140*, 504-516.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hall, T.M. (2005). Structure and function of argonaute proteins. Structure *13*, 1403-1408.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244-251.

Hock, J., Weinmann, L., Ender, C., Rudel, S., Kremmer, E., Raabe, M., Urlaub, H., and Meister, G. (2007). Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep *8*, 1052-1060.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Hutvagner, G., and Zamore, P.D. (2002). RNAi: nature abhors a double-strand. Curr Opin Genet Dev *12*, 225-232.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T., and Ishikawa, M. (2010). In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Mol Cell *39*, 282-291.

Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. Science *313*, 1134-1137.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell *39*, 292-299.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.

Jones, T., Federspiel, N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T., *et al.* (2004). The diploid genome sequence of Candida albicans. Proc Natl Acad Sci U S A *101*, 7329-7334.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467-469.

Kawamata, T., Seitz, H., and Tomari, Y. (2009). Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. Nat Struct Mol Biol *16*, 953-960.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem Sci *35*, 368-376.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kim, V.N., and Nam, J.W. (2006). Genomics of microRNA. Trends Genet *22*, 165-173.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Kufel, J., Dichtl, B., and Tollervey, D. (1999). Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. RNA *5*, 909-917.

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A *101*, 12753-12758.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol *19*, 145-151.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

Lamontagne, B., Larose, S., Boulanger, J., and Elela, S.A. (2001). The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. Curr Issues Mol Biol *3*, 71-78.

Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA *14*, 2580-2596.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol *14*, 2162-2167.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B., and MacRae, I.J. (2012). The molecular architecture of human Dicer. Nat Struct Mol Biol *19*, 436-440.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

Leulliot, N., Quevillon-Cheruel, S., Graille, M., van Tilbeurgh, H., Leeper, T.C., Godin, K.S., Edwards, T.E., Sigurdsson, S.T., Rozenkrants, N., Nagel, R.J., *et al.* (2004). A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III. EMBO J *23*, 2468-2477.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol *15*, 1501-1507.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463-2476.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol *11*, 576-577.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Liu, Y., Tan, H., Tian, H., Liang, C., Chen, S., and Liu, Q. (2011). Autoantigen La promotes efficient RNAi, antiviral response, and transposon silencing by facilitating multiple-turnover RISC catalysis. Mol Cell *44*, 502-508.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science *325*, 750-753.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.

Ma, E., MacRae, I.J., Kirsch, J.F., and Doudna, J.A. (2008). Autoinhibition of human dicer by its internal helicase domain. J Mol Biol *380*, 237-243.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

MacRae, I.J., and Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol *17*, 138-145.

MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. Nat Struct Mol Biol *14*, 934-940.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Maiti, M., Lee, H.C., and Liu, Y. (2007). QIP, a putative exonuclease, interacts with the Neurospora Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes Dev *21*, 590-600.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev *18*, 975-980.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J *12*, 861-867.

Nagel, R., and Ares, M., Jr. (2000). Substrate recognition by a eukaryotic RNase III: the double-stranded RNA-binding domain of Rnt1p selectively binds RNA containing a 5'-AGNN-3' tetraloop. RNA *6*, 1142-1156.

Nakayashiki, H., Kadotani, N., and Mayama, S. (2006). Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol *63*, 127-135.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell *121*, 1005-1016.

Nowotny, M., and Yang, W. (2006). Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. EMBO J *25*, 1924-1933.

Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

Obbard, D.J., Gordon, K.H., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci *364*, 99-115.

Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008a). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol *15*, 998.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008b). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Orban, T.I., and Izaurralde, E. (2005). Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA *11*, 459-469.

Palauqui, J.C., and Balzergue, S. (1999). Activation of systemic acquired silencing by localised introduction of DNA. Curr Biol *9*, 59-66.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol *132*, 1382-1390.

Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature *475*, 201-205.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol *12*, 1484-1495.

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Partridge, J.F., Scott, K.S., Bannister, A.J., Kouzarides, T., and Allshire, R.C. (2002). cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. Curr Biol *12*, 1652-1660.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., *et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev *18*, 2368-2379.

Pertzev, A.V., and Nicholson, A.W. (2006). Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of Escherichia coli ribonuclease III. Nucleic Acids Res *34*, 3708-3721.

Provost, P., Silverstein, R.A., Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., Wright, A., Samuelsson, B., Radmark, O., and Ekwall, K. (2002). Dicer is required for chromosome segregation and gene silencing in fission yeast cells. Proc Natl Acad Sci U S A *99*, 16648-16653.

Qin, H., Chen, F., Huan, X., Machida, S., Song, J., and Yuan, Y.A. (2010). Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. RNA *16*, 474-481.

Qu, L.H., Henras, A., Lu, Y.J., Zhou, H., Zhou, W.X., Zhu, Y.Q., Zhao, J., Henry, Y., Caizergues-Ferrer, M., and Bachellerie, J.P. (1999). Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. Mol Cell Biol *19*, 1144-1158.

Rand, T.A., Ginalski, K., Grishin, N.V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci U S A *101*, 14385-14389.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Robertson, H.D. (1982). Escherichia coli ribonuclease III cleavage sites. Cell *30*, 669-672.

Robertson, H.D., Webster, R.E., and Zinder, N.D. (1968). Purification and properties of ribonuclease III from Escherichia coli. J Biol Chem *243*, 82-91.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Rotondo, G., Huang, J.Y., and Frendewey, D. (1997). Substrate structure requirements of the Pac1 ribonuclease from Schizosaccharmyces pombe. RNA *3*, 1182-1193.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruvkun, G. (2001). Molecular biology. Glimpses of a tiny RNA world. Science *294*, 797-799.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A *104*, 8397-8402.

Schalch, T., Job, G., Noffsinger, V.J., Shanker, S., Kuscu, C., Joshua-Tor, L., and Partridge, J.F. (2009). High-affinity binding of Chp1 chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. Mol Cell *34*, 36-46.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Schwarz, D.S., Hutvagner, G., Haley, B., and Zamore, P.D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell *10*, 537-548.

Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol *14*, 787-791.

Shabalina, S.A., and Koonin, E.V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol *23*, 578-587.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.

Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244-247.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A *90*, 6498-6502.

Stoica, C., Carmichael, J.B., Parker, H., Pare, J., and Hobman, T.C. (2006). Interactions between the RNA interference effector protein Ago1 and 14-3-3 proteins: consequences for cell cycle progression. J Biol Chem *281*, 37646-37651.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc Natl Acad Sci U S A *102*, 152-157.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Tahbaz, N., Carmichael, J.B., and Hobman, T.C. (2001). GERp95 belongs to a family of signal-transducing proteins and requires Hsp90 activity for stability and Golgi localization. J Biol Chem *276*, 43294-43299.

Tahbaz, N., Kolb, F.A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T.C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO Rep *5*, 189-194.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., *et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev *17*, 49-63.

Tian, Y., Simanshu, D.K., Ascano, M., Diaz-Avalos, R., Park, A.Y., Juranek, S.A., Rice, W.J., Yin, Q., Robinson, C.V., Tuschl, T., *et al.* (2011). Multimeric assembly and biochemical characterization of the Trax-translin endonuclease complex. Nat Struct Mol Biol *18*, 658-664.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004a). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004b). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev *19*, 517-529.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev *13*, 3191-3197.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., and Stuitje, A.R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell *2*, 291-299.

Van Der Walt, J.P. (1956). Kluyveromyces- a new yeast genus of the Endomycetales. Antonie Van Leeuwenhoek *22*, 265-272.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004a). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol *14*, 346-351.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004b). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell *16*, 69-79.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Viegas, S.C., Silva, I.J., Saramago, M., Domingues, S., and Arraiano, C.M. (2011). Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway. Nucleic Acids Res *39*, 2918-2930.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wang, H.W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009a). Structural insights into RNA processing by the human RISC-loading complex. Nat Struct Mol Biol *16*, 1148-1153.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008a). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009b). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754-761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008b). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev *20*, 1732-1743.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., *et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode. Mol Cell *41*, 589-599.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol *2*, 70-75.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., *et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004). Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. Proc Natl Acad Sci U S A *101*, 8307-8312.

Wu, H., Xu, H., Miraglia, L.J., and Crooke, S.T. (2000). Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. J Biol Chem *275*, 36957-36965.

Wu, H., Yang, P.K., Butcher, S.E., Kang, S., Chanfreau, G., and Feigon, J. (2001). A novel family of RNA tetraloop structure forms the recognition site for Saccharomyces cerevisiae RNase III. EMBO J *20*, 7240-7249.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yang, D., Lu, H., and Erickson, J.W. (2000). Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in Drosophila embryos. Curr Biol *10*, 1191-1200.

Yang, W., and Steitz, T.A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. Structure *3*, 131-134.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res *34*, 667-675.

Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q., and Zhang, H. (2011). Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol *18*, 650-657.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Yin, H., and Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. Nature *450*, 304-308.

Young, R.A., and Steitz, J.A. (1978). Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in Escherichia coli ribosomal precursor RNA. Proc Natl Acad Sci U S A *75*, 3593-3597.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Zamore, P.D. (2001). Thirty-three years later, a glimpse at the ribonuclease III active site. Mol Cell *8*, 1158-1160.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zha, X., Xia, Q., and Yuan, Y.A. (2012). Structural insights into small RNA sorting and mRNA target binding by Arabidopsis Argonaute Mid domains. FEBS Lett *586*, 3200-3207.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J *21*, 5875-5885.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell *118*, 57-68.

Zhang, K., and Nicholson, A.W. (1997). Regulation of ribonuclease III processing by double-helical sequence antideterminants. Proc Natl Acad Sci U S A *94*, 13437-13441.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A *104*, 4536-4541.

Chapter 2

RNAi in Budding Yeast

Ines A. Drinnenberg[1,2,*], David E. Weinberg[1,2,3,*], Kathleen T. Xie[1,2,3,*], Jeffrey P. Mower[4,†], Kenneth H. Wolfe[4], Gerald R. Fink[1,3], David P. Bartel[1,2,3]

[1] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142
[2] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139
[3] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[4] Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland
† Present address: Center for Plant Science Innovation and Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588

* These authors contributed equally to this work.

D.E.W. performed biochemical analyses, analyzed Argonaute-associated small RNAs, generated and analyzed mRNA-seq libraries, reconstituted RNAi in *S. cerevisiae*, and demonstrated silencing of endogenous transposons. I.A.D. generated and analyzed small-RNA libraries, performed genetic analysis of *S. castellii* RNAi, and examined the phylogeny of budding-yeast Dicer. K.T.X. engineered RNAi in *S. castellii* and performed transposition assays in *S. cerevisiae*. J.P.M and K.H.W. identified the gene encoding budding-yeast Dicer. All authors designed the study and wrote the manuscript.

Published as:
Drinnenberg IA*, Weinberg DE*, Xie KT*, Mower JP, Wolfe KH, Fink GR, Bartel DP (2009). RNAi in budding yeast. *Science* 326:544–50.
* *indicates equal contributions*

RNA interference (RNAi), a gene-silencing pathway triggered by double-stranded RNA, is conserved in diverse eukaryotic species but has been lost in the model budding yeast *Saccharomyces cerevisiae.* Here, we show that RNAi is present in other budding yeast species, including *Saccharomyces castellii* and *Candida albicans*. These species use noncanonical Dicer proteins to generate small interfering RNAs, which mostly correspond to transposable elements and Y′ subtelomeric repeats. In *S. castellii*, RNAi mutants are viable but have excess Y′ messenger RNA levels. In *S. cerevisiae*, introducing Dicer and Argonaute of *S. castellii* restores RNAi, and the reconstituted pathway silences endogenous retrotransposons. These results identify a previously unknown class of Dicer proteins, bring the tool of RNAi to the study of budding yeasts, and bring the tools of budding yeast to the study of RNAi.

RNA-silencing pathways contribute to transposon silencing, viral defense, DNA elimination, heterochromatin formation, and posttranscriptional repression of cellular genes (Tomari and Zamore, 2005; Malone and Hannon, 2009). In the simplest form of silencing, known as RNA interference (RNAi), the RNaseIII endonuclease Dicer successively cleaves double-stranded RNA (dsRNA) into small interfering RNAs (siRNAs), which are loaded into the effector protein Argonaute to guide the cleavage of target transcripts (Tomari and Zamore, 2005; Farazi et al., 2008). RNAi arose in an early eukaryotic ancestor and appears to have been conserved throughout most of the fungal kingdom (Nakayashiki et al., 2006; Laurie et al., 2008) (Figure 1A). A prominent exception is *Saccharomyces cerevisiae*, a budding yeast that lacks recognizable homologs of Argonaute, Dicer, and RNA-dependent RNA polymerase (RdRP), which in some RNAi pathways produces dsRNA. Indeed, RNAi has been presumed lost in all budding yeasts. Despite this perceived loss, Argonaute genes are present in some other budding yeasts (Axelson-Fisk and Sunnerhagen, 2006; Scannell et al., 2007), including *Saccharomyces castellii* and *Kluyveromyces polysporus* (both close relatives of *S. cerevisiae*) and *Candida albicans* (the most common yeast pathogen of humans (Berman and Sudbery, 2002)) (Figure 1A). The presence of these genes in budding yeast has been enigmatic because other RNAi genes, especially Dicer, have not been found in these species. A similar conundrum appears in prokaryotes, in which certain bacteria have Argonaute homologs yet lack the other genes associated with RNAi or related RNA-silencing pathways (Hall, 2005).

siRNAs in budding yeasts

To search for RNA silencing in budding yeast, we looked for short guide RNAs, isolating 18–30-nt RNAs from *S. castellii*, *K. polysporus*, and *C. albicans* and preparing sequencing libraries representing the subset of small RNAs with 5′-monophosphates and 3′-hydroxyls (Grimson et al., 2008), which are the chemical features of Dicer products. The small RNAs of *S. castellii* and *K. polysporus* were most enriched in 23-mers beginning with U, and those of *C. albicans* were most enriched in 22-mers beginning with A or U (Figure 1B). These biases were reminiscent of those observed for Argonaute-bound guide RNAs of animals, plants, and other fungi (Lau et al., 2001; Buhler et al., 2008; Montgomery et al., 2008). Analogous RNAs were not found in *S. cerevisiae*, as expected for a species lacking RNAi (Figure 1B).

Although some reads from the Argonaute-containing yeasts mapped to ribosomal RNA (rRNA) and tRNA and presumably represented degradation intermediates of abundant RNAs, many reads clustered at other types of genomic loci. The loci generating the most reads had sequence homology to repetitive elements, including LTR retrotransposons (Ty elements), LINE-like retrotransposons (Zorro elements), and subtelomeric repeats (Y′ elements) (Figure 1C and Table S1). Loci of *S. castellii* were also particularly enriched in long inverted repeats; these palindromic loci generated most of the reads with homology to Ty elements (Figures 1C and 1D). In *S. cerevisiae*, essentially all the reads appeared to represent degradation fragments of rRNA, tRNA, and mRNA.

The reads matching inverted repeats suggested origins from paired regions of transcripts that folded back on themselves to form hairpins (Figure 1D). These inferred hairpins had 100–400-bp stems with loops ranging from 19 to >1600 nt. In regions of imperfect duplex, where reads could be mapped unambiguously, the small RNAs tended to match only one genomic strand, further supporting the idea that they originated from hairpin transcripts (Figure 1D, bottom). Other reads did not map to inverted repeats and instead mapped uniquely to both genomic strands in a pattern suggesting origins from long bimolecular duplexes involving transcripts from both strands.

Most siRNAs of the fission yeast *Schizosaccharomyces pombe* correspond to the outer repeats of the centromeres and direct heterochromatin formation and maintenance (Grewal and Jia, 2007). We therefore examined whether any of our sequenced small RNAs matched centromeres. Of the three Argonaute-containing species from which we sequenced (Figure 1B), only *C. albicans* had annotated centromeres, and almost none (<0.001%) of our *C. albicans* reads matched these genomic loci. Moreover, budding yeasts lack recognizable orthologs of the H3K9 methyltransferase Clr4 and recognizable homologs of RdRP, Tas3, Chp1, and the HP1-like chromodomain protein Swi6—proteins all necessary for RNAi-dependent heterochromatin in *S. pombe* (Grewal and Jia, 2007), arguing against a function analogous to that in *S. pombe*.

When mapped to the genome, the end of one 23-mer RNA was often next to the beginning of another 23-mer, suggesting that endonuclease cleavage simultaneously generated the 3′-terminus of one small RNA and the 5′-terminus of the next. Consistent with this hypothesis, systematic analysis of the intervals spanning the mapped ends of all 23-mer pairs revealed a clear phasing interval of 23 nt (Figure 1E). Such phasing implied successive cleavage beginning at preferred starting points. Moreover, pairs from opposite strands had the same phasing interval but in a register 2 nt offset from that of the same-strand pairs. Together, the phasing and offset implied successive cleavage of dsRNA with a 2-nt 3′ overhang—the classic biogenesis of endogenous siRNAs by Dicer (Farazi et al., 2008). Therefore, the small RNAs that appeared to derive from regions of dsRNA, i.e., those mapping in clusters to the arms of predicted hairpins and those mapping in clusters to both genomic strands, were classified as siRNAs.

Dicer in budding yeasts

The presence of siRNAs in Argonaute-containing budding yeasts implied that each of these species also had a Dicer-like activity. To assay for this activity, we monitored processing of a long dsRNA added to whole-cell extracts. Extracts from *S. castellii*, *K. polysporus*, and *C. albicans*—but not from *S. cerevisiae*—contained an activity that produced 22–23-nt RNAs, each preferentially from dsRNA rather than from single-stranded RNA (Figure 2A). Moreover, for each extract the small-RNA length matched that of the most abundant length observed *in vivo* (Figs. 1B and 2A).

Despite the observed Dicer-like activity, a gene with the domain architecture of known Dicers was not found in any budding-yeast genome (Figure 1A) (Axelson-Fisk and Sunnerhagen, 2006). Because we had evidence for cleavage of dsRNA with 2-nt 3′ overhangs, a hallmark of RNaseIII activity, we relaxed the search criteria to consider any gene with an RNaseIII domain. *S. cerevisiae* had only one gene, *RNT1*, with a recognizable RNaseIII domain. *RNT1* helps process rRNA and other noncoding RNAs (Lamontagne et al., 2001), and presumed orthologs were found throughout the fungal kingdom. *S. castellii* had a second RNaseIII-domain-containing gene, and a potential ortholog of this gene was found in each of the other Argonaute-containing budding yeasts (Figure 1A). Anticipating that this second gene encoded the Dicer of budding yeasts, we named it *DCR1*.

To test whether the Dicer candidate is required for siRNA accumulation, we deleted *DCR1* in *S. castellii*—the closest relative to *S. cerevisiae* among the sequenced Argonaute-containing species. This procedure required establishing strains and protocols to better enable molecular genetic analysis in this species (Astromskas and Cohn, 2007). In the Δ*dcr1* mutant, siRNAs failed to accumulate (Figures 2B and S2, Table S1). Deletion of the Argonaute homolog, which we named *AGO1*, also reduced siRNA accumulation, as expected if loading into Argonaute protected siRNAs from degradation (Figures 2B and S2, Table S1). For both mutants, ectopically expressing the deleted gene rescued siRNA accumulation (Figure 2B). These results indicate that the core components of endogenous RNAi pathways—Dicer, Argonaute, and siRNAs—are present in some species of the budding-yeast clade.

In *S. pombe* and other fungi, known Dicer genes resemble those in plants and animals, complete with tandem RNaseIII domains, 2–3 dsRBDs, a PAZ domain, and an N-terminal helicase domain (Bernstein et al., 2001; MacRae and Doudna, 2007) (Figure 2C). In budding yeasts, *DCR1* has two dsRBDs but only a single RNaseIII domain and no helicase or PAZ domains. Because RNaseIII domains work in pairs to nick both strands of an RNA duplex

(Zhang et al., 2004; MacRae and Doudna, 2007), we suspect that *S. castellii* Dcr1 acts as a homodimer. Dicers of insects, plants, and mammals, which already have two RNaseIII domains, do not homodimerize but do form heterodimeric complexes with cofactors that provide additional dsRBDs (Liu et al., 2003; Vazquez et al., 2004; Chendrimada et al., 2005). A homodimeric *S. castellii* Dcr1 complex would already possess four dsRBDs, which might obviate the need for such a cofactor.

Except for its second dsRBD, the domain architecture of the budding-yeast Dicer resembled that of *RNT1* rather than that of canonical Dicer genes (Figure 2C). Furthermore, the amino acid sequence of its RNaseIII domain was more similar to that of the *RNT1* RNaseIII domain than to that of any previously identified Dicer RNaseIII domain (Figure 2D). These observations suggest that budding-yeast Dicer might have emerged from a duplication of *RNT1* early in the budding-yeast lineage, perhaps coincident with the loss of canonical Dicer. The unusual ancestry and domain structure of *DCR1* might explain why its activity, and thus RNAi more generally, went undetected for so long in budding yeast.

Biochemical analyses of Dcr1 and Ago1

Dicing activity of *S. castellii* extracts was lost in the Δ*dcr1* mutant and restored by Dcr1 overexpression (Figure 2E). To determine if Dcr1 is active in the absence of *S. castellii* cofactors, we expressed the protein in *S. cerevisiae* and *E. coli* (Figure 2E). Expression in *E. coli* conferred robust activity, indicating that *S. castellii* Dcr1 is sufficient to dice dsRNA at precise intervals. In other Dicers, the PAZ domain is an essential component of a molecular ruler that imparts cleavage precision (MacRae and Doudna, 2007). The budding-yeast Dicers, which lack this domain, must achieve this measuring function differently.

To establish a biochemical link between *AGO1* and the siRNAs of *S. castellii* (Figure 2B), we sequenced the small RNAs that co-purified with tagged Ago1 expressed from its native promoter. Compared to the input RNA, the population of Ago1-associated RNAs was even more enriched for 22–23-nt RNAs and was depleted in matches to both rRNA and tRNA, with concomitant enrichment for matches to palindromes, Ty elements, and Y′ elements (Figure S3 and Table S2). These biochemical results supported the genetic link between *AGO1* and the siRNAs (Figure 2B) and provided a set of small RNAs suitable for annotating the siRNA-producing loci of *S. castellii* (Table S3).

The impact of RNAi on the *S. castellii* transcriptome

To investigate the molecular consequences of RNAi, we performed high-throughput sequencing of polyadenylated RNA (mRNA-Seq (Lister et al., 2008)) from wild-type, Δ*ago1*, and Δ*dcr1* strains (Table S4). The two annotated open reading frames (ORFs) that changed most in RNAi deletion strains were also the two with the highest density of antisense siRNA reads (Figure 3A, red points). One was the consensus Y′ ORF (Figure S5), which increased >7 fold in both deletion mutants. The other was an ORF within a palindromic Ty fragment, which increased >4 fold in the Δ*dcr1* mutant but less in the Δ*ago1* mutant. For other ORFs, transcript-abundance changes were modest and not correlated with siRNA density (Figure S6), although changes in Δ*ago1* and Δ*dcr1* mutants did correlate with each other ($R^2 = 0.39$, Figure 3A). This correlation might reflect a general response to the loss of RNAi (although we cannot exclude contributions of a common response to the hygromycin- and kanamycin-resistance genes used to delete *AGO1* and *DCR1*, respectively).

Because many siRNAs mapped antisense to or outside of ORFs, the mRNA-Seq data revealing the *S. castellii* polyadenylated transcriptome enabled the systematic identification of siRNA precursor transcripts. We focused on three types of siRNA precursors: sense-antisense transcript pairs from ORF loci, partially overlapping mRNAs, and transcripts producing the most siRNA-like reads, regardless of annotation.

The potential for dsRNA comprised of sense-antisense transcripts from ORF loci was indicated by widespread low-level antisense transcription of ORFs, with antisense mRNA-Seq tags mapping to over half of all annotated ORFs. Moreover, small RNAs mapped antisense to nearly one-third of ORFs (Figure 3A) and as a class were reduced in RNAi mutants and enriched by Ago1 immunoprecipitation (Figure S3 and Table S2). Supporting a precursor-product relationship, the abundance of the sense-antisense duplexes (inferred from mRNA-Seq data) correlated with that of small RNAs deriving from these loci (Figure S7). The most striking example of siRNAs arising from sense-antisense transcript pairs was within the Y′ ORF, which was most affected by the loss of the RNAi machinery (Figures 3A and 3B). Y′ elements are conserved protein-coding repeats. In *S. cerevisiae* they are located near both ends of most chromosomes (Louis and Haber, 1992) (Figure S7), and synteny suggests analogous locations in *S. castellii* (Byrne and Wolfe, 2005). The *S. castellii* elements had a robustly expressed antisense transcript with many siRNAs mapping to the region of sense-antisense overlap (Figure 3B).

We considered partially overlapping mRNAs as another potential source of siRNA-generating dsRNA, after using the mRNA-Seq data to extend the 5′ and 3′ boundaries of 5297 *S. castellii* protein-coding transcripts. Although only 1% of divergent transcript pairs and 7% of tandem transcript pairs overlapped, 78% of convergent transcript pairs overlapped (Figures 3C and S7). At least 43% of these convergent and overlapping gene pairs (comprising 9% of all gene pairs) generated *DCR1*-dependent siRNAs in the region of overlap (Figures 3C and S3); one such pair is illustrated (Figure 3D). A recent study reported pervasive overlapping transcripts in *S. cerevisiae* (Nagalakshmi et al., 2008). Our results revealing analogous overlap in *S. castellii* show that, in contrast to previous speculation, this phenomenon is not restricted to RNAi-deficient organisms and is an ancestral feature of these *Saccharomyces* species.

We next inferred precursor transcripts without considering whether or not they overlapped ORFs (Table S5). A hidden Markov model analyzing the Ago1-associated small RNAs identified the genomic loci producing abundant siRNAs, and analysis of the mRNA-Seq data from Δ*dcr1* strains revealed the corresponding transcripts. In addition to recovering the more prolific ORF-overlapping siRNA precursors, this analysis identified the transcript illustrated in Figure 1D and transcripts of 83 other *non-protein-coding siRNA-generating* genes of *S. castellii* (annotated as *NCS1–NCS84*, Tables S3 and S5). Transcripts producing fewer siRNAs in RNAi-competent cells changed modestly but similarly in both deletion mutants (Figure 3E), as observed when analyzing only ORF transcripts (Figure 3A). Transcripts producing the most siRNAs—which were predominantly from palindromic loci—increased dramatically in the Δ*dcr1* mutant but were relatively unchanged in the Δ*ago1* mutant (Figure 3E and Table S5), indicating that Dcr1 alone was sufficient to reduce these transcripts to wild-type levels. This mode of posttranscriptional down-regulation may be unique to palindromic transcripts, which can fold into hairpin structures that are ideal Dcr1 substrates but refractory to intermolecular pairing with Ago1-associated siRNAs.

Taken together, our results indicate that more than a thousand genomic loci in *S. castellii* generate siRNAs. The consequences of siRNAs derived from the widespread antisense and overlapping transcription in *S. castellii* are unknown. With the exception of the Y′ mRNA, the loss of the RNAi machinery did not substantially affect the levels of mRNAs corresponding to these siRNAs (Figs. 3A and S6). Perhaps in other growth conditions the regulatory impact of non-Y′ siRNAs might be more pronounced. The specificity for Y′-element regulation could arise from requiring both an abundance of antisense siRNAs and the ability to base pair with a target transcript. Although palindromic loci generate many siRNAs, the hairpin structure of these transcripts might block pairing with siRNAs, and although coding mRNAs are relatively unstructured, most generate only low levels of siRNAs. These two requirements would explain the observed impact of RNAi on the *S. castellii* transcriptome.

Engineering RNAi in *S. castellii*

To confirm that siRNAs can silence a gene in *S. castellii* and to create tools for monitoring RNAi in budding yeast, we generated two constructs (strong and weak) designed to silence a green-fluorescent-protein (GFP) reporter gene (Figure 4A). Both silencing constructs were under the control of an inducible promoter, and each was integrated into the chromosomes of wild-type, Δ*ago1*, and Δ*dcr1* strains expressing GFP. The two constructs and two induction conditions produced a gradient of *GFP* siRNAs (Figure 4B). In cells containing both *AGO1* and *DCR1*, the amount of GFP silencing, as measured by fluorescence-activated cell sorting (FACS), corresponded to the level of *GFP* siRNAs, with the highest level of siRNA production repressing fluorescence to background autofluorescence (Figure 4C). As expected, silencing depended on *DCR1* for siRNA production and on *AGO1* for siRNA function (Figures 4B and 4C). These results confirmed that siRNAs could function to silence a gene and demonstrated that the targeted transcript could originate from a locus distinct from that producing the siRNAs.

Reconstitution of RNAi in *S. cerevisiae*

Our observation that some budding yeasts closely related to *S. cerevisiae* contain a functional RNAi pathway suggested that the *S. cerevisiae* lineage lost RNAi recently and that perhaps introducing the two RNAi proteins found in *S. castellii*—Ago1 and Dcr1—could restore the pathway. To test this possibility, we used a GFP-reporter system based on our *S. castellii* system. GFP-positive strains of *S. cerevisiae* were generated that expressed either the strong, the weak, or no silencing construct. Introducing Dcr1 was sufficient to generate some *GFP* siRNAs from the weak construct and abundant *GFP* siRNAs from the strong silencing construct (Figure 4D). When Ago1 and Dcr1 were both present, we observed intermediate silencing with the weak construct and robust silencing with the strong construct (Figure 4E), with a >100-fold decrease in mRNA accompanying the decrease in fluorescence (Figures 4F and S10). Moreover, a hairpin construct targeting *URA3* reduced growth in the absence of uracil and enabled growth on 5-fluoroorotic acid (5-FOA), demonstrating that the RNAi pathway reconstituted in *S. cerevisiae* can silence an endogenous gene with phenotypic consequences (Figure 4G).

The ability to reconstitute RNAi in *S. cerevisiae* using only Ago1 and Dcr1 raises the possibility that the *S. castellii* RNAi pathway requires only these two proteins. This simplicity would make budding-yeast RNAi distinct from all known RNAi pathways, which use additional proteins involved in, for example, Argonaute loading (e.g., R2D2 in *Drosophila melanogaster* (Tomari et al., 2004)) or maturation of the silencing complex (e.g., QIP in *Neurospora crassa* (Maiti et al., 2007)). The four dsRBDs that would be present in a Dcr1 homodimer might explain the absence of a separate loading factor. Alternatively, overexpression of Ago1, Dcr1, and a hairpin precursor might be sufficient to enact RNAi in *S. cerevisiae*, but they might require additional factors for efficient silencing when expressed at physiological levels in *S. castellii*.

Another possibility is that the reconstituted pathway uses components that have been maintained in *S. cerevisiae* since its recent loss of RNAi.

RNAi and transposon silencing

The Δ*ago1* and Δ*dcr1* mutants of *S. castellii* were viable, with no obvious growth disadvantage in minimal or rich media at a range of temperatures, no observed decrease in mating, sporulation, or chromosome stability, and no altered sensitivity to a replication inhibitor (hydroxyurea) or to microtubule destabilizing agents (thiobendazole and benomyl). However, both Δ*ago1* and Δ*dcr1* mutants had difficulty retaining introduced plasmids, demonstrating that the loss of RNAi has detectable phenotypic consequences (Figure S11).

We suspected that budding-yeast RNAi might also silence transposable elements. RNAi and related processes silence and eliminate transposons in other eukaryotes (Malone and Hannon, 2009), and a large fraction of our budding-yeast siRNAs corresponded to transposable elements. For example, most *S. castellii* siRNAs mapped to fragments of Ty retrotransposons (Figure 1C). Despite the abundance of Ty fragments, indicative of former activity in the *S. castellii* lineage (Figure S12), we have not yet found an active retrotransposon in the current, albeit incomplete, *S. castellii* genome sequence. Therefore, to test the effect of RNAi on transposition, we turned to the RNAi-competent *S. cerevisiae* strain.

Compared to the strain with no RNAi genes or the one with only *DCR1*, the RNAi-competent strain had much less Ty1 Gag protein and mRNA (Figures 5A and 5B). The dsRNA triggering this repression of protein and mRNA from native Ty1 elements could have come from elements expressing their own antisense transcripts (Berretta et al., 2008) or from neighboring elements or fragments oriented with potential to produce convergent or hairpin transcripts (Kim et al., 1998). Analysis of published mRNA-Seq data from *S. cerevisiae* (Ingolia et al., 2009) revealed regions with many tags antisense to Ty1 elements (Figure 5C), and these regions produced siRNAs in *S. cerevisiae* strains containing *DCR1* (Figure 5D). To examine whether RNAi can suppress retrotransposition, we ectopically expressed a Ty1 element marked with *HIS3*, which enabled transposition to be detected as plasmid-independent complementation of histidine auxotrophy (Garfinkel et al., 1988). Consistent with our molecular findings for endogenous elements, the RNAi-competent strain permitted much less transposition (Figure 5E). These results, combined with our sequencing data (Figure 1C), indicate that a major role of budding-yeast RNAi is to silence transposons.

Adding the minimal RNAi components conferred transposon silencing to a species normally lacking the RNAi pathway. The recipient strain had no obvious abnormalities while endogenous transposon protein and mRNA were both drastically reduced, which illustrates the ability of RNAi to preferentially target transposon genes rather than other cellular genes. While specific for transposable elements, the pathway appears general for any element requiring an RNA transcript—including those it had not previously encountered—by exploiting internally initiated antisense transcripts as well as the intrinsic propensity of these elements to generate hairpin and convergent transcripts as their genomic load increases.

Concluding remarks

We have uncovered an RNAi pathway present in several different budding-yeast species that appears distinct from the well-characterized pathway of fission yeast. The two known components of the pathway have a patchy phylogenetic distribution among budding yeasts (Figure 1A), indicating that the pathway can be lost easily. Indeed, if transposon silencing is the critical function of the RNAi pathway, then a species in which transposons have been completely silenced for a long evolutionary period is likely to lose all intact elements and thereby lose selection to retain the RNAi pathway, opening the door to re-invasion. Perhaps also contributing to RNAi loss is its potential inhibition of dsRNA viruses and their associated satellite dsRNAs. In *S. cerevisiae*, the M satellite element of the reovirus-like L-A virus encodes a secreted toxin that kills neighboring cells lacking element-encoded immunity (Wickner, 1996). If cells that have lost RNAi are better able to retain this system, they might have a selective advantage despite having lost an efficient transposon-defense pathway.

With the discovery and characterization of the budding-yeast pathway, RNAi can be used as a tool to silence genes in *S. cerevisiae*, *S. castellii*, and presumably other budding yeasts. RNAi might be particularly useful in *C. albicans*, an obligate diploid for which both gene deletions and genetic screens are not trivial (Berman and Sudbery, 2002). Even in *S. cerevisiae*, RNAi might have advantages for repressing repetitive gene families. RNAi also enables an inducible repression system that might provide an alternative to existing technologies, which involve either non-physiological expression of the gene of interest (e.g. the GAL/GLU system) or generation of temperature-sensitive mutations. Perhaps more importantly, the tools of budding yeast can now be applied to the study of RNAi, either by developing reagents to investigate the endogenous pathway in *S. castellii* or by applying existing technologies to examine the reconstituted pathway in *S. cerevisiae*. While anticipating a productive future for RNAi research in budding yeasts, we note that if in the past *S. castellii* rather than *S. cerevisiae* had been chosen as the model budding yeast, the history of RNAi research would have been dramatically different.

Acknowledgements

We thank A. Hochwagen, G. Ruby, and H. Guo for helpful discussions; V. Agarwal for protein homology modeling; D. Garfinkel for Ty1-VLP antiserum and other reagents; M. Cohn and A. Regev for yeast strains; W. Johnston and the Whitehead Genome Technology Core for high-throughput sequencing; and S. Pääbo for academic advising. Supported by Science Foundation Ireland and the Irish Research Council for Science, Engineering, and Technology (K.H.W.), NIH grants GM040266 (G.R.F.), GM0305010 (G.R.F.), and GM067031 (D.P.B.), an NSF graduate research fellowship (D.E.W.), and a Boehringer-Ingelheim Fonds predoctoral fellowship (I.A.D.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

Accession numbers

The Gene Expression Omnibus accession number for all the mRNA-Seq and small-RNA cloning data is GSE17872.

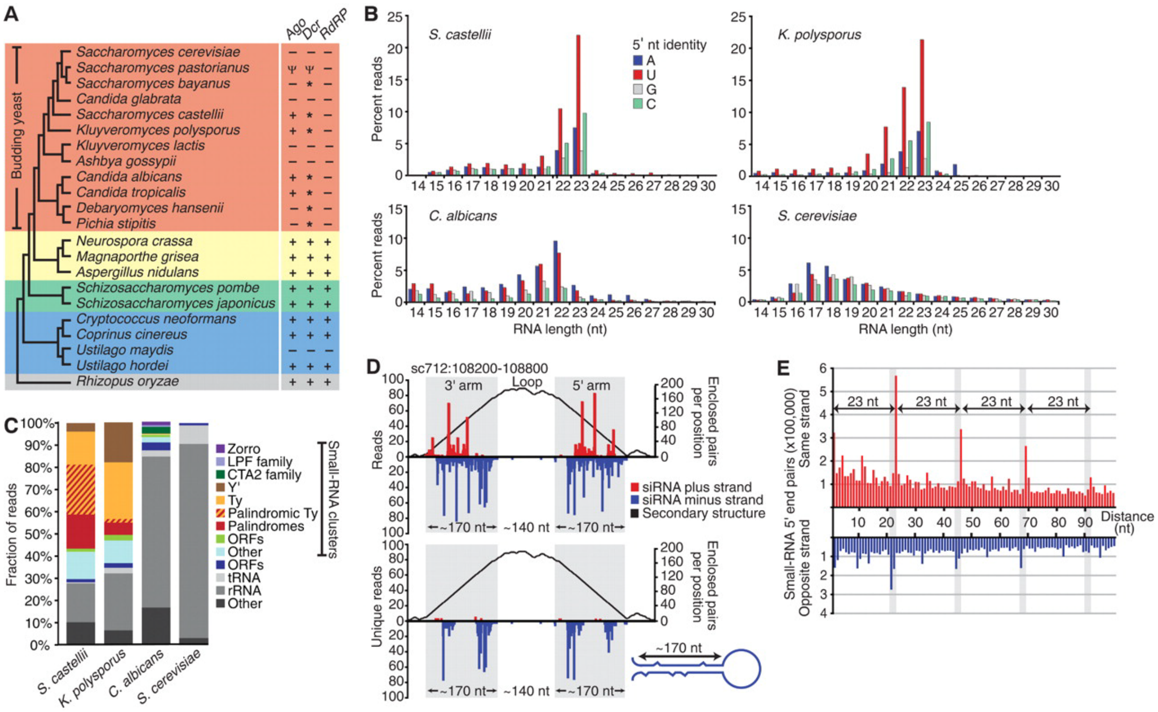


Figure 1. Endogenous siRNAs in budding yeasts

(A) Cladogram showing Basidiomycota (blue), Zygomycota (gray) and Ascomycota, subdivided into Saccharomycotina (budding yeasts, orange), Pezizomycotina (yellow), and Taphrinomycotina (green) (Fitzpatrick et al., 2006; Hedtke et al., 2006). The presence of canonical RNAi genes is indicated (+) [(Nakayashiki et al., 2006; Laurie et al., 2008); and references therein]. All genomes had an *RNT1* ortholog, and several others had a second RNaseIII domain–containing gene (*), which has Dicer activity in *S. castellii*. Pseudogenes are indicated (Ψ). *S. bayanus*, which had a Dicer but not an Argonaute gene, appeared to lack siRNAs (Figure S1).

(B) Length distribution of genome-matching sequencing reads representing small RNAs with the indicated 5′ nucleotide. Reads matching rRNA and tRNA are excluded.

(C) Classification of loci to which 21- to 23-nt RNAs map, considering those that map to clusters in a pattern suggestive of siRNAs separately from those that do not.

(D) A palindromic region generating siRNAs in *S. castellii*. 5′ termini of 22- to 23-nt RNAs were mapped to the genome, and counts (normalized to the number of genomic matches) are plotted for the plus and minus genomic strands. The top considers all reads; the bottom considers those matching the genome at only one locus. The predicted structure of the (–)-strand transcript is represented as a mountain plot (Zuker and Stiegler, 1981).

(E) Distribution of the genomic intervals separating the 5′ termini of sequenced 23-nt RNAs from *S. castellii*. Plotted is the frequency of each interval, when considering all pairs of reads less than 100 nt apart (excluding reads matching rRNA and tRNA).

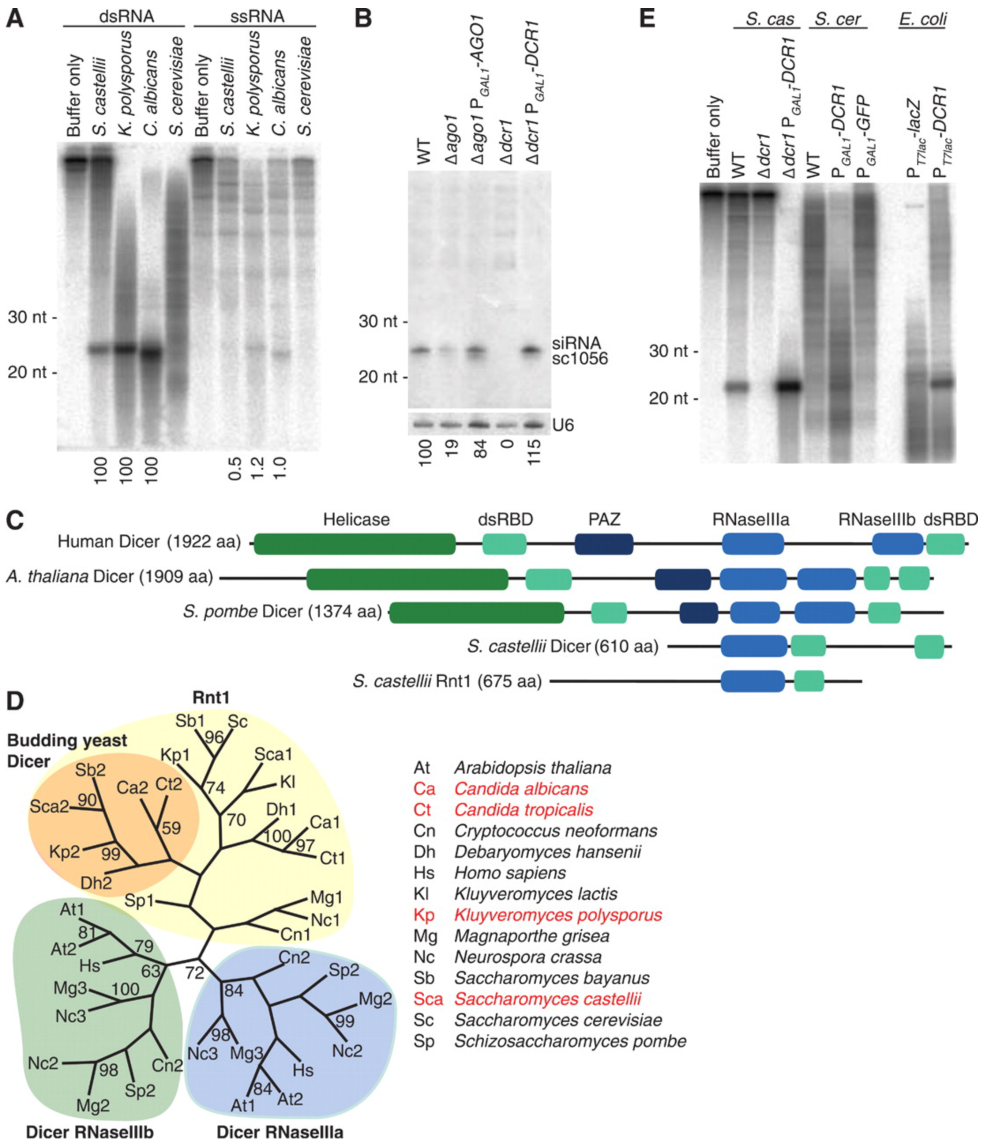


Figure 2. The Dicer of budding yeast

(A) In vitro processing of radiolabeled dsRNA or single-stranded RNA (ssRNA) in extracts from the indicated budding yeast species. Products were resolved on a denaturing gel. The fraction of product normalized to that observed with dsRNA is indicated below as a percentage.

(B) RNA blot probing for an endogenous siRNA (sc1056) in the indicated deletion and rescue strains. The blot was reprobed for U6 small nuclear RNA, and the siRNA percent signal normalized to that of U6 is indicated below.

(C) Domain architectures of representative Dicer proteins and the two *S. castellii* proteins containing an RNaseIII domain.

(D) Maximum-likelihood tree based on amino acid alignment of RNaseIII domains from Dicer proteins and Rnt1 homologs. Orange shading highlights budding yeast Dicer candidates indicated by asterisks in Figure 1A. Budding yeast species encoding Argonaute are listed in red. Bootstrap values higher than 50% are shown.

(E) In vitro dicing in extracts from recombinant *S. castellii* (*S. cas*), *S. cerevisiae* (*S. cer*), or *E. coli* strains with the indicated deletions and additions, analyzed as in (A).

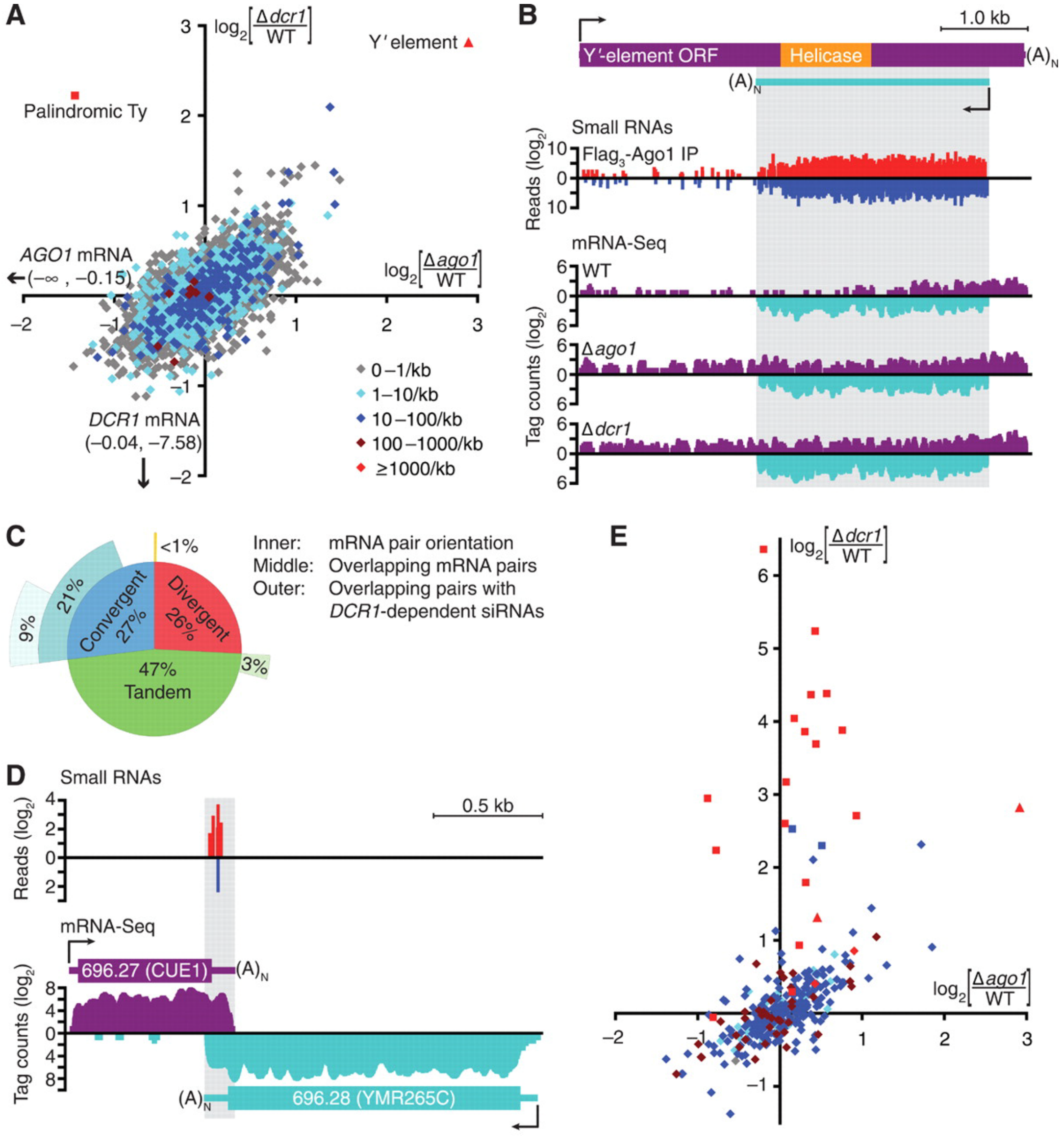


Figure 3. The impact of RNAi on the *S. castellii* transcriptome

(A) Strand-specific mRNA-Seq analysis of annotated ORF transcripts in wild-type (WT) and RNAi-mutant strains. Plotted is the $\log_2$ ratio of transcript abundance in Δ*ago1* versus wild-type (*x* axis) and Δ*dcr1* versus wild-type (*y* axis). Colors indicate the density (reads per kilobase) of antisense small (22- to 23-nt) RNAs that copurified with Ago1. A Ty ORF fragment (annotated as Scas_712.50) embedded within a palindromic siRNA-producing locus is indicated (square). Annotated Y′-element ORFs were replaced by one consensus Y′ ORF (triangle, Figure S5). Because the mRNA-Seq protocol included poly(A) selection, which retains the 3′ but not 5′ fragments of cleaved mRNAs, we calculated full-length transcript abundance using tags mapping to the 5′ half of each ORF. Similar trends were observed when we used tags mapping across the ORF (Figure S4).

(B) Analysis of the *S. castellii* Y′ element. The numbers of siRNA 5′ ends (small RNAs) and mRNA tags (mRNA-Seq) mapping to the consensus Y′ element are plotted for each position (sense, above axis; antisense, below axis).

(C) Gene-pair organization and overlap in *S. castellii*. (Inner ring) Fraction of neighboring annotated ORFs with the indicated orientation; (middle ring) fraction of transcript pairs with overlapping 3′ ends (convergent), overlapping 5′ ends (divergent), or continuous transcription in between (tandem); (outer ring) fraction of convergent transcript pairs generating siRNAs in the overlapping region.
(D) A pair of convergent transcripts that generate siRNAs in the region of overlap. Plots are as in (B).
(E) mRNA-Seq analysis of inferred siRNA-generating transcripts. The plot is as in (A), with the same colors to indicate siRNA-read density and shapes to indicate transcripts mapping to Y′ elements (triangle), palindromes (square), and others (diamonds).

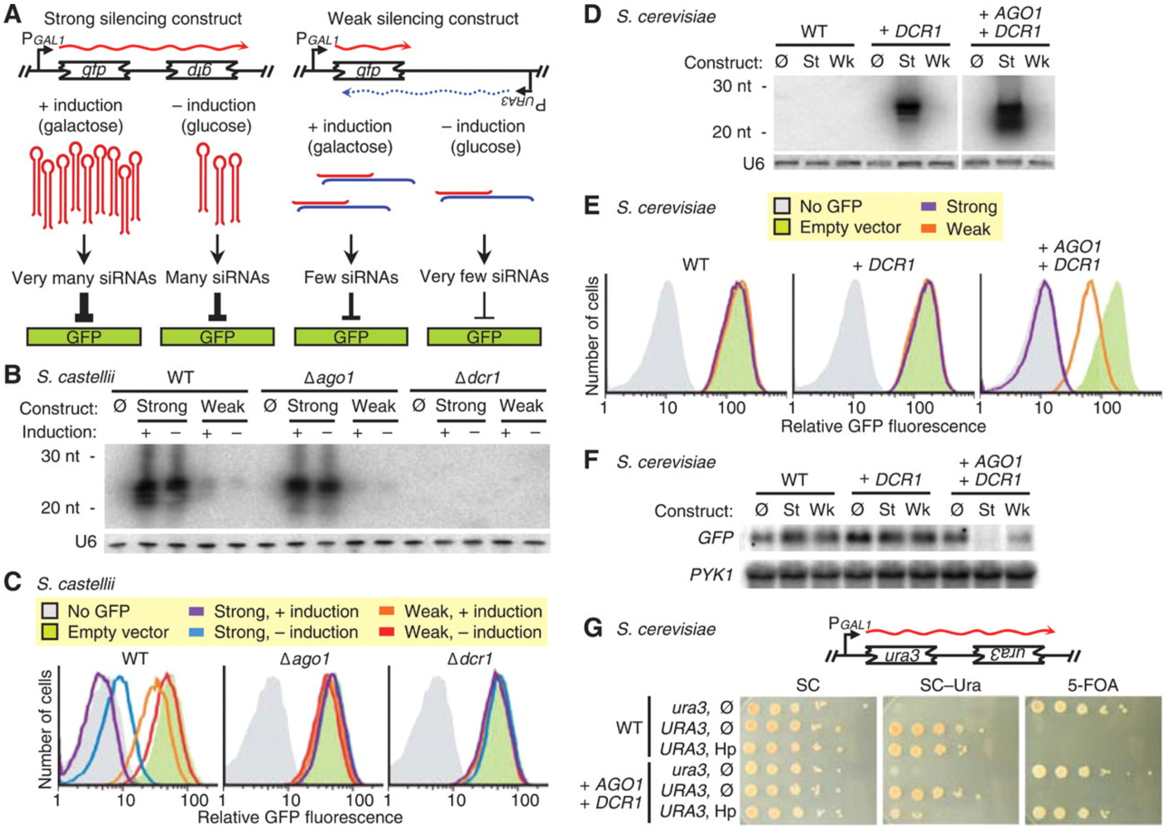


Figure 4. Engineering RNAi in *S. castellii* and *S. cerevisiae*

(A) Schematic for silencing of a GFP reporter. The strong silencing construct included inverted repeats of a *gfp* fragment and was designed to produce a hairpin transcript (Sigova et al., 2004). The weak silencing construct contained one copy of the fragment, which is transcribed convergently to produce dsRNA.

(B) RNA blot probing for siRNAs antisense to *GFP*, by using total RNA from the indicated *S. castellii* strains with integrated empty vector (Ø) or silencing construct (strong or weak), either induced with galactose (+) or uninduced (–). The blot was reprobed for U6 small nuclear RNA.

(C) FACS histograms showing GFP fluorescence in the indicated *S. castellii* strains expressing the indicated silencing constructs.

(D) RNA blot probing for siRNAs antisense to *GFP* in *S. cerevisiae* strains expressing either no *S. castellii* genes (WT) or the indicated integrated *S. castellii* genes, and either the strong (St), the weak (Wk), or no (Ø) silencing construct. The blot was reprobed for U6 small nuclear RNA.

(E) FACS histograms showing GFP fluorescence in the indicated *S. cerevisiae* strains expressing the indicated silencing constructs. All strains were induced; silencing from uninduced constructs was similar for the strong construct and undetectable for the weak construct (Figure S9).

(F) RNA blot probing for *GFP* mRNA in the indicated *S. cerevisiae* strains expressing the indicated silencing constructs. The blot was reprobed for *PYK1* mRNA as a loading control.

(G) Silencing an endogenous gene. *S. cerevisiae* strains containing nonfunctional and functional *URA3* genes (*ura3* and *URA3*, respectively) and expressing the indicated *S. castellii* genes and either the diagrammed hairpin construct (Hp) or no silencing construct (Ø) were tested for Ura3p expression by plating serial dilutions on complete medium (SC), medium lacking uracil (SC–Ura), and medium containing 5-FOA (to which cells producing Ura3p are sensitive).

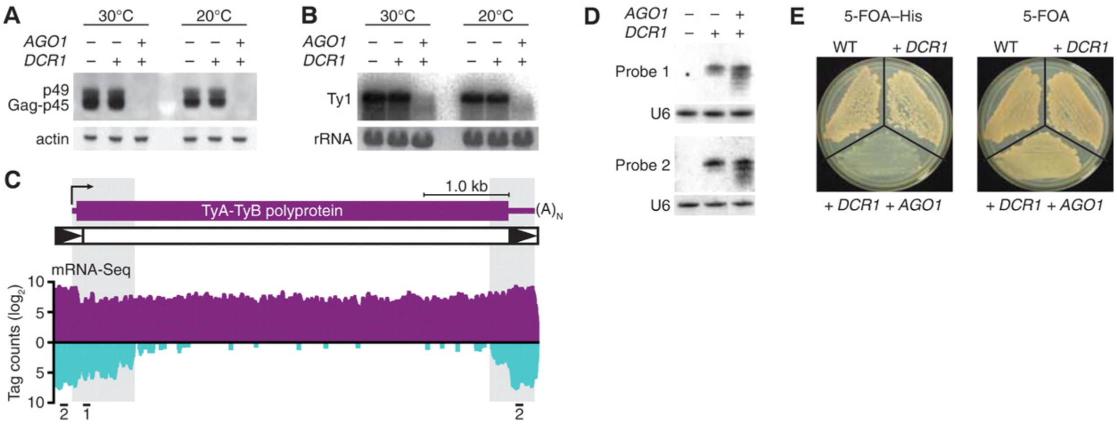


Figure 5. Silencing of Ty1 retrotransposons by RNAi in *S. cerevisiae*
(A) Immunoblot probing for Ty1 Gag protein (p45) and its precursor (p49) (Kawakami et al., 1993) in *S. cerevisiae* strains expressing the indicated *S. castellii* genes. Strains were grown under standard (30°C) or transposition-inducing (20°C) conditions. The blot was reprobed for actin.
(B) RNA blot probing for Ty1 mRNA, analyzing the same cultures as in (A). Ethidium bromide–stained rRNA is shown.
(C) mRNA-Seq analysis of *S. cerevisiae* Ty1 elements. The average numbers of mRNA tags to a consensus Ty1 element are plotted (sense, above axis; antisense, below axis). The schematic shows a Ty1 transcript (purple) and element, with long terminal repeats as black triangles. Locations of the two probes used in (D) are indicated.
(D) RNA blot probing for siRNAs processed from endogenous Ty1 dsRNA. The blot was reprobed for U6 small nuclear RNA.
(E) *HIS3*-marked Ty1 transposition assay. Galactose-induced *S. cerevisiae* strains expressing the indicated *S. castellii* genes were tested for transposition by growth on plates with 5-FOA and lacking histidine (5-FOA–His). Cells grow without histidine and are resistant to 5-FOA when the *HIS3*-marked Ty1 element has transposed into the genome and the *URA3*-marked plasmid carrying the original *HIS3*-marked element has been lost (Garfinkel et al., 1988). Also shown is growth on media selective for plasmid loss but not transposition (5-FOA).

References

Astromskas, E., and Cohn, M. (2007). Tools and methods for genetic analysis of Saccharomyces castellii. Yeast *24*, 499-509.

Axelson-Fisk, M., and Sunnerhagen, P. (2006). Comparative genomics and gene finding in fungi. Comparative Genomics: Using Fungi as Models *15*, 1-28.

Berman, J., and Sudbery, P.E. (2002). Candida Albicans: a molecular revolution built on lessons from budding yeast. Nat Rev Genet *3*, 918-930.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev *22*, 615-626.

Buhler, M., Spies, N., Bartel, D.P., and Moazed, D. (2008). TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. Nat Struct Mol Biol *15*, 1015-1023.

Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res *15*, 1456-1461.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Farazi, T.A., Juranek, S.A., and Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development *135*, 1201-1214.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., and Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol *6*, 99.

Garfinkel, D.J., Mastrangelo, M.F., Sanders, N.J., Shafer, B.K., and Strathern, J.N. (1988). Transposon tagging using Ty elements in yeast. Genetics *120*, 95-108.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35-46.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Hall, T.M. (2005). Structure and function of argonaute proteins. Structure *13*, 1403-1408.

Hedtke, S.M., Townsend, T.M., and Hillis, D.M. (2006). Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst Biol *55*, 522-529.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Kawakami, K., Pande, S., Faiola, B., Moore, D.P., Boeke, J.D., Farabaugh, P.J., Strathern, J.N., Nakamura, Y., and Garfinkel, D.J. (1993). A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics *135*, 309-320.

Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., and Voytas, D.F. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res *8*, 464-478.

Lamontagne, B., Larose, S., Boulanger, J., and Elela, S.A. (2001). The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. Curr Issues Mol Biol *3*, 71-78.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Laurie, J.D., Linning, R., and Bakkeren, G. (2008). Hallmarks of RNA silencing are found in the smut fungus Ustilago hordei but not in its close relative Ustilago maydis. Curr Genet *53*, 49-58.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Louis, E.J., and Haber, J.E. (1992). The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevisiae. Genetics *131*, 559-574.

MacRae, I.J., and Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol *17*, 138-145.

Maiti, M., Lee, H.C., and Liu, Y. (2007). QIP, a putative exonuclease, interacts with the Neurospora Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes Dev *21*, 590-600.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell *133*, 128-141.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Nakayashiki, H., Kadotani, N., and Mayama, S. (2006). Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol *63*, 127-135.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A *104*, 8397-8402.

Sigova, A., Rhind, N., and Zamore, P.D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe. Genes Dev *18*, 2359-2367.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev *19*, 517-529.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol *14*, 346-351.

Wickner, R.B. (1996). Double-stranded RNA viruses of Saccharomyces cerevisiae. Microbiol Rev *60*, 250-265.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell *118*, 57-68.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res *9*, 133-148.

Drinnenberg et al., *Science* 2009

Supporting Online Material

Materials and Methods

Growth conditions and genetic manipulations

S. castellii was grown at 25°C on standard *S. cerevisiae* plate and liquid media (e.g., YPD and SC). Transformations were performed as described (*S1*) with some modifications. Either 0.5–2 µg plasmid DNA or 1–7 µg linear DNA was added to 5 µl single-stranded DNA (10 mg/ml salmon sperm DNA, Sigma D7656), mixed with 50 µl yeast (~3 x 10^8 cells in 100 mM lithium acetate), and added to transformation buffer (a mixture of 240 µl 40% PEG 3350 and 36 µl 1 M lithium acetate). After incubation at 25°C for 30–90 min, 35 µl of DMSO was added, and the entire mixture was incubated at 42°C for 10 min, resuspended, and then plated on selective media.

Other species. Growth temperatures were as follows, unless otherwise noted: *K. polysporus*, 25°C; *S. cerevisiae*, *S. bayanus*, and *C. albicans*, 30°C; *E. coli*, 37°C.

Strain construction

A list of strains used and generated in this study is provided (table S7).

Heterothallic strains. Most of our strains started with the homothallic *S. castellii* strain Y235 (*ura3-1*/*ura3-1*, *Ho*/*Ho*), generously provided by M. Cohn (*ura3-1* is a point mutation G541A that creates the amino acid substitution G181R). To delete the Ho endonuclease, the loxP-KanMX6-loxP module of plasmid pUG6 (*S2*) was used as a template to amplify the disruption cassette by fusion PCR (*S3*), with ~400-bp targeting arms on both sides of the cassette (primers 5′-TGATCGAAGAAGGCACTAGAA and 5′-CAGATCCACTAGTGGCCTATGCGGCCGCTGTCATTGAAAATCGCCAAA, 5′-GCGTACGAAGCTTCAGCTGGCGGCCGCGGCCAAATTCTTCCTGCAACT and 5′-TTTTCGGACTTCACGAGCTT). The resulting heterozygous strain (*ura3-1*/*ura3-1*, *Ho*/*ho::loxP-KanMX-loxP*) was transformed with pSH47 (*S2*), which encodes the *Cre* recombinase under the control of the *S. cerevisiae GAL1* promoter. The expression of *Cre* was induced for 2 h in liquid culture, and strains sensitive to G418 were isolated. This strain was transferred to sporulation medium (1% potassium acetate, 0.1% yeast extract, 0.05% glucose) for 4 days, and tetrads were dissected. Although sporulation efficiency and spore viability were generally low in Y235, stable heterothallic strains of mating type **a** and α (DPB004 and DPB005, respectively) could be derived from a tetrad with four viable spores, showing that *S. castellii ho* deletion strains could not switch mating type.

Deletion of AGO1 *and* DCR1. *AGO1* and *DCR1* were deleted using the hygromycin cassette of pAG32 (*S4*) and the loxP-KanMX6-loxP cassette of pUG6 as dominant selection markers, respectively. For diploids, homozygous deletions (DPB002 and DPB003) were generated first by deleting one copy in Y235, sporulating the resulting heterozygotes, and allowing isolated spores to grow, switch mating types, and mate. *AGO1* and *DCR1* were deleted in DPB004 and DPB005 to generate DPB006, DPB007, DPB008, DPB009, and DPB313. The *AGO1* disruption construct was created as follows: *AGO1* was amplified from genomic DNA (5′-TGAACGTGTGGAAGACCAAA and

5′-AGTGGCTAACGGCAACATATCAGACA) and cloned into pCR4Blunt-TOPO (Invitrogen); the hygromycin cassette was then inserted between the *Hind*III and *Age*I restriction sites within the *AGO1* genomic fragment; the *AGO1* disruption construct was then amplified with the same primers used for *AGO1* cloning. Deletion of *DCR1* was analogous to deletion of *Ho* (fusion PCR primers 5′-TTCAACACCTCCAGCAACAG and 5′-CAGATCCACTAGTGGCCTATGCGGCCGCAGGCATTGCAACAATCTGTG, 5′-GCGTACGAAGCTTCAGCTGGCGGCCGCGCTGTTGCTGGAGGTGTTGAA and 5′-TTTACCACCATACCATGAGTTTTT).

Tagged Ago1 strain for immunoprecipitation. A haploid strain expressing Flag$_3$-tagged Ago1 from its native promoter (DPB220) was constructed by two-step homologous recombination in DPB005, as follows: a *S. cerevisiae URA3* expression cassette (amplified from pYES2.1, Invitrogen) was used to replace the start codon of *AGO1* by transformation and selection of transformants on SC–ura plates; the *URA3* cassette was subsequently replaced by a Flag$_3$ tag (amplified with a start codon from pQCXIP, gift of D. Sabatini) by transformation and selection on 5-FOA.

S. castellii *GFP reporter strains*. The loxP-KanMX6-loxP cassette in DPB009 was removed by *Cre* expression as described above to generate DPB318. The GFP(S65T)-KanMX6 module from pFA6a (*S5*) was then integrated at the endogenous *ura3* locus in DPB005, DPB313, and DPB318 (such that GFP was fused in-frame directly after the ATG start codon of *ura3*) to generate GFP-expressing strains DPB314, DPB317, and DPB321. The silencing constructs (pIp, pIp-weakSC_GFP, and pIp-strongSC_GFP) were integrated upstream of the ORF annotated as Scas_633.2 in DPB314, DPB317, and DPB321 to create strains DPB331–DPB339. For these integrations, each silencing construct was linearized by digestion with *Sac*I, and 1.5 µg was transformed. Transformants were selected on SC–ura plates.

S. cerevisiae *RNAi reporter strains*. The GFP(S65T)-KanMX6 module from pFA6a was integrated at the endogenous *ura3* locus in L4718 to create DPB249. Integration of Ago1 and Dcr1 expression vectors (pRS404-P$_{TEF}$-Ago1 and pRS405-P$_{TEF}$-Dcr1) and GFP silencing construct vectors (pRS403-P$_{GAL1}$-weakSC_GFP and pRS403-P$_{GAL1}$-strongSC_GFP) into the genome was done by linearization and transformation using standard protocols (*S6*) to create DPB250, DPB251, and DPB255–DPB260. To generate strains useful for *URA3* silencing, DPB249 and DPB258 were transformed with functional *URA3* coding sequence amplified from pRS406 to create the uracil prototrophs DPB271 and DPB275, respectively. Integration of the silencing construct pRS403-P$_{GAL1}$-hpSC_URA3 into DPB271 and DPB275 generated DPB272 and DPB276, respectively.

Plasmid construction

A list of plasmids generated in this study is provided (table S8).

Yeast Ago1 *and* Dcr1 *expression plasmids*. *S. castellii AGO1* or *DCR1* was cloned into pYES2.1 (Invitrogen) to produce the galactose-inducible Ago1 and Dcr1 expression plasmids pYES2.1-Ago1 and pYES2.1-Dcr1, respectively. *GFP* was also cloned into pYES2.1 (creating pYES2.1-GFP) as a negative control.

E. coli *recombinant expression plasmids*. For recombinant expression of Dcr1 in *E. coli*, *DCR1* was cloned into pET101/D-TOPO, creating pET101-Dcr1. pET101-lacZ was supplied by the manufacturer (Invitrogen).

S. castellii *GFP silencing constructs*. A multiple cloning site containing *Xho*I and *Eco*RI restriction sites was cloned between the *Pvu*II and *Xba*I restriction sites of pYES2.1. For the strong silencing construct, 275 bp of *GFP* sequence from pFA6a was then cloned in the sense orientation between *Pvu*II and *Xho*I sites, and in the antisense orientation between *Eco*RI and *Xba*I sites, in *E. coli* SURE (Stratagene). The weak silencing construct was made identically, except without *GFP* sequence in the antisense orientation. A 73-bp sequence spanning intron 1 from *S. pombe rad9* was then added between *Xho*I and *Eco*RI sites (modeled after (*S7*)). To convert these episomal plasmids into integrating plasmids, the 2-micron and f1 origins were then replaced (using *Nhe*I and *Spe*I sites) by sequence from *S. castellii* sc633:288301–289016 (amplified from genomic DNA with 5′-`AAAAGCTAGCGATCCCTTATCAAATATGGTAC` and 5′-`AAAAACTAGTGTAGAATCCAGAGAATAGAATC`). These resulting *S. castellii* integrating plasmids expressing weak and strong *GFP* silencing constructs are pIp-weakSC_GFP and pIp-strongSC_GFP, respectively. The pIp empty vector was created by replacing the hairpin of pIp-strongSC_GFP with *Xho*I and *Eag*I sites.

S. cerevisiae *reconstitution and silencing constructs*. Vectors pRS404-P_{TEF}-Ago1 and pRS405-P_{TEF}-Dcr1 were constructed by insertion of the coding sequence of the respective *S. castellii* genes between the *TEF* promoter and *CYC1* terminator (cloned from pRS416-P_{TEF} (*S8*)) of the appropriate vector (*S9*) using *Spe*I and *Xho*I sites (Ago1) or *Xba*I and *Xho*I sites (Dcr1). To generate vectors pRS403-P_{GAL1}-strongSC_GFP and pRS403-P_{GAL1}-weakSC_GFP, an expression cassette containing the *GAL1* promoter, *CYC1* terminator, and GFP silencing construct sequence was cloned out of the appropriate episomal pYES2.1 silencing construct into the *Not*I and *Sal*I sites of pRS403. To generate the *URA3* silencing vectors, 339 bp of *URA3* sequence from pRS406 was initially cloned into the episomal pYES2.1 GFP weak silencing construct in the sense orientation between *Pvu*II and *Xho*I sites (thereby replacing the *GFP* sequence), and in the antisense orientation between *Eco*RI and *Xba*I sites. pRS403-P_{GAL1}-hpSC_URA3 was then created by cloning an expression cassette containing the *GAL1* promoter, *CYC1* terminator, and *URA3* silencing construct sequence out of this pYES2.1 plasmid into the *Not*I and *Sal*I sites of pRS403.

In vitro dsRNase assays

Substrates. Blunt-ended dsRNA substrate was prepared by simultaneous *in vitro* transcription from two PCR templates carrying T7 promoter sequences at opposite ends. Reactions were assembled using the MAXIscript Kit (Ambion) with a 32:1 molar ratio of UTP:[α-^{32}P]UTP (800 Ci/mmol) according to the manufacturer's directions. Control ssRNA was prepared similarly, except that a single PCR template was included in the transcription reaction. DNase-treated RNA was fractionated on a 4% urea gel, eluted from gel slices in 0.3 M NaCl overnight at 4°C, and ethanol precipitated.

Strains. Wild-type strains used in Figure 2A were *S. castellii* Y235, *K. polysporus* KpolWT, *C. albicans* Can14, and *S. cerevisiae* FY45. Strains used in Figure 2E were as follows: *S. castellii*, DPB005, DPB318, and DPB318 transformed with pYES2.1-Dcr1; *S. cerevisiae*, F2005 and F2005 transformed with either pYES2.1-Dcr1 or pYES2.1-GFP; *E. coli*, BL21 Star(DE3) (Invitrogen) transformed with either pET101-lacZ or pET101-Dcr1.

Extracts. Strains in Figure 2A were grown in YPD to OD_{600} 1.2–1.6; yeast strains in Figure 2E were grown similarly, except P_{GAL1} strains were grown in SC–ura with galactose/raffinose, and all strains were grown at 25°C; *E. coli* were grown in LB with 100 µg/ml ampicillin to OD_{600} 0.6 and induced (1 mM IPTG) for 4 h. Cells were harvested by centrifugation and flash frozen in 100–200 mg aliquots. Aliquots were thawed on ice, resuspended in 1 µl/mg lysis buffer [50 mM HEPES pH 7.6, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM EGTA, 300 mM sodium acetate, 5% glycerol, 0.25% NP-40, protease inhibitor cocktail (Roche), 1 mM PMSF], and vortexed four times for 45 s at 4°C with an equal volume of glass beads. Lysates were clarified by centrifugation at 10,000x *g* for 5 min. Extract concentrations were normalized according to absorbance at 260 nm and stored at –80°C.

Reactions. The 20 µl reactions contained 10 µl extract (or 10 µl lysis buffer for "Buffer only" control), 4 µl 5X reaction buffer (125 mM HEPES pH 7.2, 10 mM magnesium acetate, 10 mM DTT, 5 mM ATP), and 10,000 cpm radiolabeled substrate. In Figure 2A, reactions were incubated at 25°C (*S. castellii* and *K. polysporus*) or 30°C (all others) for 2 h; in Figure 2E all reactions were incubated at 25°C. Reactions were quenched with AE Buffer (50 mM sodium acetate pH 5.5, 10 mM EDTA) and phenol extracted.

RNA blots

Total RNA was isolated using the hot phenol method. Small RNA blots were performed using 10–15 µg total RNA per lane and carbodiimide-mediated cross-linking to the membrane (*S10*), with the following DNA probes radiolabeled at their 5′ termini: *S. castellii* siRNA sc1056, 5′-CTATCTTCATCGATTACCATCTA; *S. castellii* U6 small nuclear RNA, 5′-TATGCAGGGGAACTGCTGAT; *GFP* siRNA, 5′-ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCA; Ty1 probe 1, 5′-CCGTTAGACGTTTCAGCTTCCAAAACAGAAGAATGTGAGAAGGCTTCCACTAAG; Ty1 probe 2, 5′-TAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAAC. mRNA blots were performed using 4–5 µg DNase-treated total RNA per lane and UV crosslinking. *GFP* and Ty1 body-labeled antisense riboprobes were prepared by using PCR products as templates for in vitro transcription (MaxiScript kit, Ambion). A radiolabeled *PYK1* (*CDC19*) DNA probe was prepared by random priming (Prime-It II, Stratagene).

Strains used in Figure 2B were Y235, DPB002, DPB002 transformed with pYES2.1-Ago1, DPB003, and DPB003 transformed with pYES2.1-Dcr1. Strains used in Figure 4B were DPB331–DPB339. Strains used in Figure 4D and 4F were DPB249–DPB251, and DPB255–DPB260. Strains used in Figure 5D were DPB249, DPB255, and DPB258.

RT-PCR

Reverse transcription reactions were performed with 100 ng total RNA using Superscript III according to the manufacturer's instructions (Invitrogen) with the following gene-specific primers in the same reaction: *GFP*, 5′-TGTGGTCTCTCTTTTCGTTGG; *ACT1*, 5′-TCAAAGAAGCCAAGATAGAACCA. PCR reactions were assembled in 100 µl with 2 µl RT reaction using the following primers: *GFP*, 5′-TTTCACTGGAGTTGTCCCAAT and 5′-GAAAGGGCAGATTGTGTGG; *ACT1*, 5′-ACGTTGGTGATGAAGCTCAA and 5′-ATACCTGGGAACATGGTGGT. After the indicated number of cycles, a 15 µl aliquot was removed and combined with 3 µl 6X DNA loading dye. 6 µl was loaded onto a 1.5% agarose gel, and DNA was visualized by EtBr staining.

Plasmid loss

DPB005, DPB313, and DPB008 were transformed with 1.5 µg pRS316 (*S8*), pYES2.1-weakSC_GFP, pYES2.1-Ago1, or pYES2.1-Dcr1. Transformants were plated directly on SC–ura plates containing 2% glucose (uninduced) or 2% galactose (induced). To analyze plasmid loss, cells from colonies were inoculated in 5 ml of the medium indicated in Figure S11 and passaged once a day for 4 days.

Southern blots

Each lane contained 2 µg of RNA-free DNA isolated as described in (*S11*) and digested with *Xba*I. Plasmids were detected using a probe with the ampicillin-resistance gene sequence (amplified using primers `5′-CCATGAGTGATAACACTGCG` and `5′-GGCACCTATCTCAGCGATC`). The genomic locus was detected using a probe with sequence from *S. castellii* sc718:138001–138427 (amplified using primers `5′-GCATAAGCTGTGCTTTAGACT` and `5′-CTTGTAACGGTTCAATTCTAGC`).

FACS analysis

Two biological replicates of each strain were inoculated in SC, either noninducing (2% glucose) or inducing (*S. castellii*, 2% galactose; *S. cerevisiae*, 1% galactose and 1% raffinose), and grown overnight. Fresh cultures were then seeded from the overnight cultures and cells were grown to log-phase. Cells were analyzed using FACSCalibur (BD Biosciences); data were processed with CellQuest Pro (BD Biosciences) and FlowJo (Tree Star). The same samples were used for RNA and GFP analyses.

S. cerevisiae URA3 silencing

Strains (DPB249, DPB271–DPB272, DPB258, DPB275–DPB276) were inoculated in SC under inducing conditions (1% galactose and 1% raffinose) and grown for 1 day. Cells were diluted to OD_{600} of 1.0, and 1:10 serial dilutions were spotted onto the appropriate plates (SC, SC–ura, or 5-FOA; all containing 1% galactose and 1% raffinose) and grown at 30°C for 3 days.

S. cerevisiae Ty1 analysis

Transposition assay. Strains (DPB249, DPB255, and DPB258) were transformed with 1 µg of pGTy*H3HIS3* (galactose-inducible Ty1 marked with *HIS3*, where transcription of the Ty1 and *HIS3* is in the same direction) (*S12*) and selected on SC–ura plates. Transformants were streaked out on SC–ura with 2% galactose plates and grown at 20°C for 2 days to induce transposition. Cells were then replica-plated onto YPD plates and grown at 30°C for 1 day for plasmid loss. These cells were then replica-plated onto 5-FOA–his plates (to select for both plasmid loss and transposition) or 5-FOA plates (to select for plasmid loss only) and grown at 30°C for 2–3 days.

When using the more standard *his3*-artificial-intron marker for retrotransposition (*S13*), analogous results were obtained but were not as informative because the marker produces a non-physiological antisense transcript, which could pair with the sense transcript to generate an ectopic dsRNA trigger.

RNA and protein analysis. Strains (DPB249, DPB255, and DPB258) were inoculated in SC and grown overnight. For non-transposition-inducing conditions, cells were diluted to OD 0.125 and grown at 30°C to OD_{600} 0.9–1.0. For transposition-inducing conditions, cells were diluted to 100 cells/ml and grown at 20°C to OD_{600} 0.9–1.0. Cells were harvested by centrifugation and flash frozen.

Immunoblotting. Three OD_{600} units of cells were resuspended in 100 ml H_2O. After adding 160 µl of extraction buffer (1.85 M NaOH, 7.4% β-mercaptoethanol), cells were incubated on ice for 10 min. 160 µl of 50% trichloroacetic acid was added, and cells were incubated on ice for an additional 10 min. Precipitated material was collected by centrifugation, and the supernatant was discarded. The tube was washed with 500 µl of 1 M Tris pH 8.0, centrifuged briefly, and the supernatant was discarded. The pellet was vigorously resuspended in 150 µl of 1X Laemlli sample buffer and boiled for 4 min. Samples (12 µl each) were resolved by SDS-PAGE, transferred to poly(vinylidene difluoride) in CAPS-ethanol pH 10, and probed sequentially with Ty1-VLP antiserum (*S14, 15*) and anti-actin (Abcam, ab8224). Immunoblots were developed with HRP-conjugated anti-rabbit or anti-mouse antibodies and enhanced chemiluminescence (Amersham).

Small-RNA sequencing and analysis

Library preparation. Total RNA was isolated using hot phenol from log-phase YPD cultures of *S. castellii* F2037, *K. polysporus* KpolWT, *S. cerevisiae* FY45, *S. bayanus* F2035, and *C. albicans* Can14. Small-RNA cDNA libraries were prepared as described (*S16*) and sequenced using the Illumina SBS platform. Libraries were also prepared and sequenced from RNAi deletion strains (DPB002 and DPB003).

Ago1 immunoprecipitation. A saturated overnight culture of DPB249 was diluted to OD_{600} 0.3 in 150 ml YPD and grown to OD_{600} 1.5. Extracts were prepared as for *in vitro* dsRNase assays. For the input fraction, one-fifth of the extract was removed and added to AE buffer. Anti-Flag M2 agarose (Sigma) was incubated with the remaining extract at 4°C for 1.5 h. Beads were washed with lysis buffer four times, after which the remaining buffer was removed and AE buffer was added. Small RNA libraries were prepared as described above.

Read processing. After removing the adaptor sequences, reads representing the small RNAs were collapsed to a non-redundant set, and 14–30-nt sequences were mapped to the appropriate genome, allowing no mismatches and recovering all hits (table S1). When counting the reads matching a locus, the count was hit-normalized, i.e., normalized to the number of times that a small-RNA sequence matched the genome. For example, a small RNA sequenced twice that mapped to the genome five times contributed 0.4 read counts to each genomic locus. Sequence and feature files for *S. cerevisiae* S288C and *C. albicans* SC5314 were obtained from the *Saccharomyces* Genome Database (SGD) on September 10, 2007 and the *Candida* Genome Database Assembly 21. Sequence files for *S. bayanus* MCYC623 that were current as of January 18, 2009 were downloaded from NCBI. Sequence and feature files for *S. castellii* CBS4309 and

K. polysporus DSM70294 were obtained from the Yeast Gene Order Browser (YGOB) (*S17*). Using the set of *S. cerevisiae* tRNA and rRNA sequences as queries for blastn alignments (e-value cutoff, e-10), genomic loci mapping to tRNA and rRNA in *S. castellii*, *K. polysporus*, and *S. bayanus* were identified. In *K. polysporus*, tRNA and rRNA annotations were available in the GenBank flatfile obtained from YGOB and used to supplement the alignments.

Initial identification of siRNA clusters. For the small RNAs sequenced from total RNA, genomic regions giving rise to siRNAs were identified by parsing the genome files from *S. castellii*, *K. polysporus*, and *C. albicans* into non-overlapping windows of 500 bp. Windows with high levels of siRNA expression (22–23-nt sequences for *S. castellii* and *K. polysporus*, 21–22-nt sequences for *C. albicans*; excluding tRNA and rRNA reads) were selected by applying read and sequence density cutoffs manually adjusted based on the data set (*S. castellii*, ≥10 reads/kb or ≥10 genome matches/kb; *K. polysporus*, ≥50 reads/kb or ≥50 genome matches/kb; *C. albicans*, ≥40 reads/kb or ≥40 genome matches/kb). Adjacent windows passing the density cutoffs were concatenated. The small-RNA profile of each of these clusters was manually inspected for adherence to properties, including length (23 nt for *S. castellii* and *K. polysporus*; 22 nt for *C. albicans*) and 5′-nt biases (U for *S. castellii* and *K. polysporus*; A or U for *C. albicans*).

Refined identification of siRNA clusters in S. castellii. Using sequencing reads of small RNAs co-purifying with Ago1, a hidden Markov model (HMM) was constructed with two states, "C" (giving rise to siRNAs) and "N" (not giving rise to siRNAs). The ratio of 23-mer reads relative to all reads (excluding 22-mer reads) was calculated in 10-bp windows (apportioning hit-normalized counts to the windows based on the fraction of its nucleotides covered by the small RNA) to define two types of emissions: 0) ratio ≥0.45 and 1) ratio <0.45. Emission probabilities were generated by training on the initially identified siRNA clusters to represent the "C" state, and training on five supercontigs (sc1014, sc621, sc542, sc534 and sc587) to represent the "N" state. Transition probabilities for the given window size were estimated using the median length of these siRNA clusters (250 bp) that map to Y′ elements and palindromic arms, or the average length of the intervening genomic sequence between two clusters, i.e. the difference derived from the total length of all contigs (11,354,548 bp) divided by the number of clusters identified in the initial analysis (100). Initial state probabilities were calculated based on the proportion of contigs in "C" state, i.e. total length of siRNA clusters (25,000 bp) divided by the total length of all contigs. Using the Viterbi algorithm, the contigs were parsed over non-overlapping 10-bp windows. The parse yielded 379 clusters (table S3) with the three regions that map to rRNA excluded. The cluster boundaries were adjusted to include the full sequence of all small RNAs with at least one nucleotide mapping to the cluster and to exclude terminal nucleotides not covered by a small RNA.

Cluster annotation. Clusters were further characterized based on previous genome annotations and alignments. Reads for Figure 1C (21–23 nt) and for figure S3 (22–23 nt) were classified into categories. Reads of siRNA clusters that mapped to annotated ORFs in either sense or antisense orientation were grouped together in Figure 1C as reads from ORF clusters. Using the Flag$_3$-Ago1 IP dataset, siRNA reads in clusters overlapping ORFs were further separated into "clusters sense to ORF" and "clusters antisense to ORF." siRNA reads that mapped to convergent overlapping ORF transcripts (annotated using the mRNA-Seq dataset) were categorized as "overlap."

The DNA sequences of the siRNA clusters from the *S. castellii* and *K. polysporus* datasets were aligned against the *S. cerevisiae* protein dataset (NCBI) using blastx (e-value cutoff 0.001). Significant alignments to Ty elements were extended 300 nt on both sides, and reads overlapping these extended alignments were classified as Ty-proximal siRNA reads. Additional Ty elements could be identified using annotated Ty elements from (*S18*) as blastx queries. More careful Ty annotations for *S. castellii* could then be made by identifying *S. castellii* Ty LTR, gag, and pol sequences based on the initial blastx matches and other Ty sequence signatures ((*S18-20*) and references therein). Similarly, siRNA clusters derived from Y′ elements were detected. For cases in which siRNA expression exceeded the boundaries of the annotated Y′ element ORF in a processive, un-gapped fashion, those siRNAs were still classified as Y′-element-proximal siRNAs. siRNA clusters in *C. albicans* were annotated based on the *C. albicans* genome annotation and blastx alignments against the set of protein sequences downloaded from NCBI (e-value cutoff 0.001).

Palindromes were predicted using the IRF program (*S21*) with the following parameters: Alignment Parameters, 2, 3, and 5 (match, mismatch, and indels, respectively); minimum Alignment Score To Report Repeat, 100; T4 small palindromes (20–80+ nt) loop length, 100 nt; T5 medium palindromes (80–300+ nt) loop length, 1000 nt; T7 large palindromes (300–2400+ nt) loop length, 5000 nt; maximal loop length, 5000 nt; maximal stem length, 10,000 nt; allow GT matches. The following numbers of palindromes were identified: 66 in *S. castellii*, 222 in *K. polysporus*, 61 in *C. albicans*, and 390 in *S. cerevisiae*. These palindromes were compared to our lists of siRNA-generating loci. In most cases when overlap was observed, the 22–23-nt RNAs were enriched in the inverted-repeat regions rather than the intervening region or surrounding regions. In some cases the palindromes overlapped with each other and the one with 22–23-nt RNAs mapping to the repeats was the one chosen. In some cases (10 of 43 for *S. castellii*, and 42 of 90 for *K. polysporus*), the overlap of 22–23-nt RNAs was not preferentially at the repeats; these were not classified as palindromic clusters. Using the initial datasets, these analyses revealed 19 palindromic siRNA clusters in *S. castellii* and 29 in *K. polysporus*, all of which either overlapped or were contained within the set of siRNA clusters identified by the sliding window approach. The refined cluster identification based on the Flag$_3$-Ago1 IP dataset from *S. castellii* revealed 23 palindromic siRNA clusters (table S5).

Phasing analysis. The frequency of distances separating all 23-mer 5′-end pairs (*i*, *j*) mapping to the same DNA strand was calculated using the following equation:

$$Frequency_D = \sum_{i,j} (Reads_i \cdot Reads_j)_D$$

where D = distance between sRNA 5′ ends

The frequency of distances separating pairs of 23-mer 5′ ends mapping to opposite strands of DNA was calculated separately using the same equation.

Phylogenetic analysis

Psi symbols for *S. pastorianus* (Fig. 1A) indicate a highly degraded *AGO1* pseudogene and a *DCR1* pseudogene that is intact except for a single internal stop codon. The intact *S. bayanus DCR1* gene shows conservation of amino acid sequence relative to the *S. pastorianus* pseudogene (*dN*/*dS* ratio 0.3) despite the absence of intact *AGO1* in both species. The *AGO1* and *DCR1* loci are syntenic among *S. castellii*, *K. polysporus*, *S. pastorianus,* and *S. bayanus*.

A maximum-likelihood (ML) tree of RNaseIII domains was constructed using the PHYLIP software package (http://evolution.genetics.washington.edu/phylip.html). RNaseIII domains were predicted using SMART (*S22, 23*). The amino acid sequences of the RNaseIII domains were used to compute a multiple sequence alignment using TCOFFEE (*S24*). A consensus ML tree was built by running DNAML (PHYLIP) on the amino acid alignment after bootstrap re-sampling (500 replicates) of the data set using SEQBOOT (PHYLIP). The phylogenetic tree was displayed using TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

Protein name/accession numbers used in Figure 2D are as follows: At1, *A. thalania* DCL1; At2, *A. thalania* DCL2; Ca1, *C. albicans* EAK98282; Ca2, *C. albicans* XP_717277; Ct1, *C. tropicalis* AAFN01000070; Ct2, *C. tropicalis* AAFN01000057; Cn1, *C. neoformans* XP_569593.1; Cn2, *C. neoformans* XP_569797.1; Dh1, *D. hansenii* XP_457483.1; Dh2, *D. hansenii* XP_457193.1; Hs, *H. sapiens* DICER1; Kl, *K. lactis* F2416; Kp1, *K. polysporus* 1045p1; Kp2, *K. polysporus* 455p11; Mg1, *M. grisiae* XP_363615; Mg2, *M. grisiae* XP_367242; Mg3, *M. grisiae* XP_367242; Nc1, *N. crassa* Sms3; Nc2, *N. crassa* Dcl2; Nc3, *N. crassa* NCU01762; Sb1, *S. bayanus* 671p65; Sb2, *S. bayanus* 643p2; Sca1, *S. castellii* 696p6; Sca2, *S. castellii* 626p5; Sc, *S. cerevisiae* Rnt1; Sp1, *S. pombe* Pac1; Sp2, *S. pombe* Dcr1.

To search for a PAZ domain in the *S. pombe* Dcr1 protein, the full-length protein sequence was submitted as a query to the HHpred server, allowing 10 PSI-BLAST iterations and 1000 maximum hits (*S25*). All available standard HMM databases (pdb70_3Sep09, pdb_on_hold_3Sep09, scop70_1.71, scop70_1.75, cdd_17Jul09, interpro_16.2, pfamA_23.0, smart_17Jul09, panther_4Mar08, tigrfam_4Mar08, pirsf_4Mar08, COG_17Jul09, KOG_17Jul09, CATH_4Mar08, supfam_4Mar08, pfam_17Jul09, cd_17Jul09, test56, test18) were searched, and the results realigned with the Maximum ACcuracy (MAC) alignment algorithm. The search retrieved a family that included full-length Dicers (KOG id: KOG0701; E-value = 0) and a crystallized Dicer derived from *Giardia intestinalis* (PDB id: 2qvw/2ffl; E-value = 1.1e-15), which both aligned to the *S. pombe* Dcr1 in regions that included their PAZ domains. The search also retrieved known PAZ-domain entries (CDD ids: cd02843, cd02844, cd02845, KOG id: KOG1042, Pfam id: PF02170; E-values = 7.9e-05, 2.3e-05, 0.95, 19, 0.018, respectively), even though these entries lacked flanking domains to aid in the alignment. The same procedure was performed replacing the full-length Dcr1 query with a region between the first dsRNA-binding domain and the first RNaseIII domain (a.a. 628–914 of the *S. pombe* Dcr1 protein), which encompassed the putative PAZ region but no other known domains. This search also retrieved Dicer proteins and other known PAZ domain entries. Analogous searches did not provide evidence for a PAZ domain in *S. castelli* Dcr1p.

mRNA sequencing and analysis

Strand-specific mRNA-Seq. Two biological replicates of DPB005 (WT), DPB007 (Δ*ago1*), and DPB009 (Δ*dcr1*) were grown in YPD to OD_{600} 0.6–0.8. Total RNA isolated using hot phenol was treated with DNaseI (RiboPure-Yeast Kit, Ambion). Poly-$(A)^+$ mRNA was purified from 75 μg total RNA using magnetic oligo-dT DynaBeads (Invitrogen) according to the manufacturer's instructions, and then fragmented by alkaline hydrolysis (*S26*). Trace amounts of synthetic 3′-

pCp[5′-^{32}P]-labeled 26-nt and 32-nt RNA size markers were added to monitor the subsequent steps. RNA fragments (25–45 nt) were gel-purified and 3′-dephosphorylated in a 25 µl reaction containing 12.5 units T4 PNK (New England Biolabs) and MES-NaOH buffer (100 mM MES-NaOH pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 300 mM NaCl) for 6 h at 37°C. After phenol extraction and precipitation, RNA was ligated to pre-adenylated adaptor DNA as described (*S16*). Gel-purified ligation products were 5′-phosphorylated in a 14 µl reaction containing 15 units T4 PNK and PNK buffer for 30 min at 37°C. After phenol extraction and precipitation, RNA was ligated to a 5′-adaptor RNA, gel-purified, converted to cDNA, amplified, and sequenced as described (*S16*).

Read processing. The first 25 nt of each 36-nt read were isolated and collapsed into a non-redundant set of 25-nt sequences with occurrence counts (table S4). Sequences were mapped to the reference genome, allowing no mismatches and recovering all hits. Transcript-specific analysis of small-RNA data (e.g., Fig. 3A) was based on 22–23-nt reads from the $Flag_3$-Ago1 IP dataset, unless indicated otherwise.

Exon annotations were downloaded from YGOB (introns less than 10 nt were considered sequencing errors and assigned as exons). Sense mRNA, antisense mRNA, and antisense small-RNA read counts were calculated individually for each gene by summing the hit-normalized reads mapping either to the 5′-half of the ORF (mRNA tags, half-ORF analysis) or across all of the ORF (small-RNA reads); a sequence contributed $N \cdot nt/25$ reads to a gene (N = hit-normalized read number; nt = number of nt in the 25-nt sequence overlapping the ORF). In parallel, mRNA tag counts were also calculated across the entire ORF (full-ORF analysis, fig. S4).

For each gene, mRNA-Seq tag counts from biological replicates were averaged. Genes for which none of the three strains had an average tag count above 20 (half-ORF analysis) or above 30 (full-ORF analysis), and ORFs corresponding to Y′ element fragments, were excluded from all analyses except in figures S4A and S4B. mRNA abundance was calculated by dividing tag counts by kb of mapped exon. mRNA-Seq tag counts from Δ*ago1* were normalized to those of WT by first ranking genes based on the ratio of tags in Δ*ago1* versus WT, and then multiplying the WT tag counts by a factor such that the median ranked gene had a transcript abundance ratio of 1. An analogous normalization procedure was also applied to Δ*dcr1*. The final normalization factors were 0.8847 for WT, 1.0000 for Δ*ago1*, and 0.8440 for Δ*dcr1*. The same normalization factors were applied to the single-nucleotide-resolution mRNA-Seq plots for the Y′ element consensus (Fig. 3B).

Consensus Y′ element of S. castellii. An initial set of Y′-element fragments was obtained by extending and combining annotated Y′-element ORFs and Y′-element fragments manually identified in the course of annotating siRNA clusters. These fragments were assembled into a single contig using SeqMan Pro (DNASTAR Lasergene). The resulting majority sequence was used as a query for blastn against the genome (e-value cutoff 10^{-10}, MegaBlast option). All additional Y′ element fragments obtained from this search were added to the consensus, bringing to 32 the total number of unique contributing genomic fragments (fig. S5). mRNA tags and small-RNA reads were mapped to the consensus Y′ element independently of the genome. Each library was initially mapped to the set of Y′ element fragments, allowing no mismatches and recovering all hits. Mapped nucleotide positions with respect to fragments were converted into positions with respect to the consensus. Mapping data was normalized using the above factors and used to generate single-nucleotide-resolution plots of the consensus Y′ element (Fig. 3B).

Y′ element transcript and siRNA abundances were the sum of read and tag nucleotides across the region of interest divided by the appropriate length (25 nt for mRNA; 22 or 23 nt for siRNA).

Comparing ORF-derived siRNA levels with transcript levels. For each annotated protein-coding gene, mRNA tags and small-RNA reads mapping across its ORF were determined as above, except only uniquely mapping sequences were included. For each ORF, sense and antisense transcript abundances were estimated separately as the sum of tags from all six mRNA-Seq libraries (without normalization), and siRNA abundance was estimated as the sum of sense and antisense small-RNA reads. Genes with no unique mRNA-Seq tags mapping to the coding strand were excluded. Genes were ranked by total transcript abundance (sum of sense and antisense tags) and by inferred duplex abundance (minimum of sense and antisense tags). Genes with non-zero abundance were divided into three equally sized bins (high, mid, low). For inferred duplex analysis, genes with zero inferred duplex abundance (i.e., genes with sense tags but no antisense tags) formed a fourth bin.

Transcripts corresponding to siRNA-generating loci. For each siRNA cluster identified using the HMM, two transcripts—one on each strand—were initiated and assigned the coordinates of the cluster. Tags from Δ*dcr1* mRNA-Seq libraries were used to extend cluster transcripts as follows. The transcript was extended 10 nt in the 5′ direction if that 10-nt window had a tag density within 10-fold (above or below) of that of the initially assigned transcript. This process was iterated using the average tag density of the extended transcript. Once a window failing this criterion was reached, the transcript was terminated before the window. Then, the 3′ end was also thus extended, beginning with the average tag density of the transcript that included the extended 5′ end. Transcript extension was also tried first in the 3′ then in the 5′ direction; when the transcript ends disagreed between these two orders, the combination of 5′ and 3′ ends forming the largest transcript was used. The ends were then more finely mapped by identifying the first nucleotide upstream and last nucleotide downstream that corresponded to any tags (in Δ*dcr1* mRNA-Seq libraries), with a maximum extension of 10 additional nucleotides. Coordinates of inferred transcripts are presented in table S3. Transcripts that had mRNA-Seq tags mapping to them but that did not overlap any previous annotations were annotated as *non-coding-siRNA-generating* genes (*NCS*, table S3).

Transcript abundance in each mRNA-Seq library and siRNA abundance were determined as with coding transcripts, with the following exceptions: intron annotations were ignored, and an average read cutoff of 15 tags (half-transcript analysis) or 20 tags (full-transcript analysis) in any strain was applied. Y′-element fragments were removed and replaced with the consensus, except in table S3.

Protein-coding transcript extension and overlap. Of 5693 annotated ORFs, 5297 (93%) had mRNA-Seq tags mapping to at least 70% of the ORF nucleotides (combining tags from all three strains) and were carried forward for further analysis. For each ORF, the 5′ and 3′ boundaries of the transcript were extended using the mRNA-Seq tags, requiring contiguous tag coverage outward from the ORF boundaries and assigning the revised 5′ and 3′ boundaries to the most distal nucleotides represented by these mRNA-Seq tags.

A gene pair was defined as a gene and its right neighbor (according to YGOB annotations). The 5297 ORFs were parsed into 4776 gene pairs, with the loss of pairs attributable mainly to genes located at the ends of contigs. The number of convergent overlapping transcripts giving rise to *DCR1*-dependent siRNAs was calculated comparing 22–23-nt reads from the Flag$_3$-Ago1 input and Δ*dcr1* datasets. 467 convergent overlapping loci had uniquely mapping small RNA reads in the Flag_3-Ago1 input dataset. The Δ*dcr1* dataset was then used to adjust this number to account for the loci for which small RNAs represented *DCR1*-independent mRNA degradation intermediates. Because RNA degradation intermediates would be overrepresented in the Δ*dcr1* small RNA dataset due to the absence of siRNAs, the Δ*dcr1* dataset was normalized to the Flag_3-Ago1 input dataset based on the number of rRNA and tRNA reads. Three normalized Δ*dcr1* datasets were constructed from the complete dataset by random sampling without replacement. In these three datasets, a median of 30 convergent overlapping loci had uniquely mapping Δ*dcr1* small RNA reads, which indicated that at least 437 convergent overlapping loci (43%) gave rise to *DCR1*-dependent uniquely mapping siRNAs.

To compare overlapping transcripts between *S. castellii* and *S. cerevisiae*, a list of gene pairs with opposite and overlapping transcripts in *S. cerevisiae* was downloaded from http://www.yale.edu/snyder/ (*S27*). The genes comprising these 828 unique gene pairs were mapped to their corresponding *S. castellii* genes based on YGOB annotations. 398 pairs corresponded to annotated convergent gene pairs in *S. castellii*. These pairs were cross-referenced with the list of *S. castellii* overlapping convergent gene pairs to determine the number producing overlapping transcripts in both species. Of the convergent gene pairs syntenic between these two genomes and reported to generate overlapping mRNAs in *S. cerevisiae* (*S27*), 84% generated overlapping mRNAs in *S. castellii*.

S. cerevisiae *mRNA-Seq analysis.* Strand-specific mRNA-Seq data from *S. cerevisiae* (*S26*) was downloaded from the Gene Expression Omnibus (samples GSM346117 and GSM346118) and processed as for *S. castellii*. Telomere annotations (TEL16L, TEL16R, TEL12L, and TEL12R) were downloaded from SGD, and hit-normalized tag counts were used to plot the abundance of mRNA-Seq tags at single-nucleotide resolution (i.e. tags contributed to counts along their entire length). To analyze mRNA-Seq tags mapping to a consensus Ty1 element, the 28 full-length Ty1 elements in the S288C genome sequence (identified using Ty1*H3* as a query for blastn against the genome) were aligned using SeqMan Pro (DNASTAR Lasergene). mRNA-Seq tags matching the consensus element were analyzed as for the consensus Y′ element of *S. castellii*, except tag counts were divided by 28 to obtain the average number of tags per full-length element at each position.

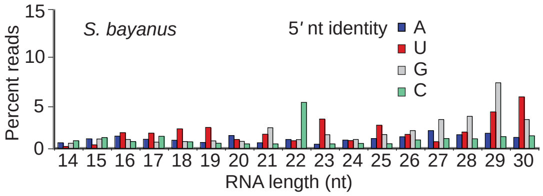


Figure S1. Analysis of small-RNA library from *S. bayanus* MCYC623. Length distribution of genome-matching reads (as percent of reads that do not match tRNA or rRNA) representing small RNAs with the indicated 5′ nucleotide (nt). Reads matching tRNAs and rRNAs were excluded.

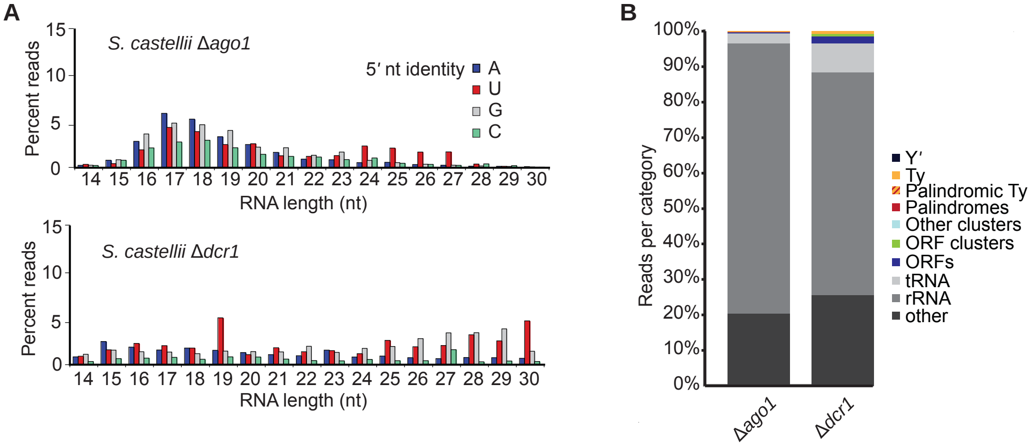


Figure S2. Analysis of small-RNA libraries from RNAi-mutant strains. **(A)** Length distributions of genome-matching reads (as percent of reads that do not match tRNA or rRNA) representing small RNAs with the indicated 5′ nucleotide (nt). Reads matching tRNAs and rRNAs were excluded.
(B) Classification of 21–23-nt reads based on genome annotations and alignments.

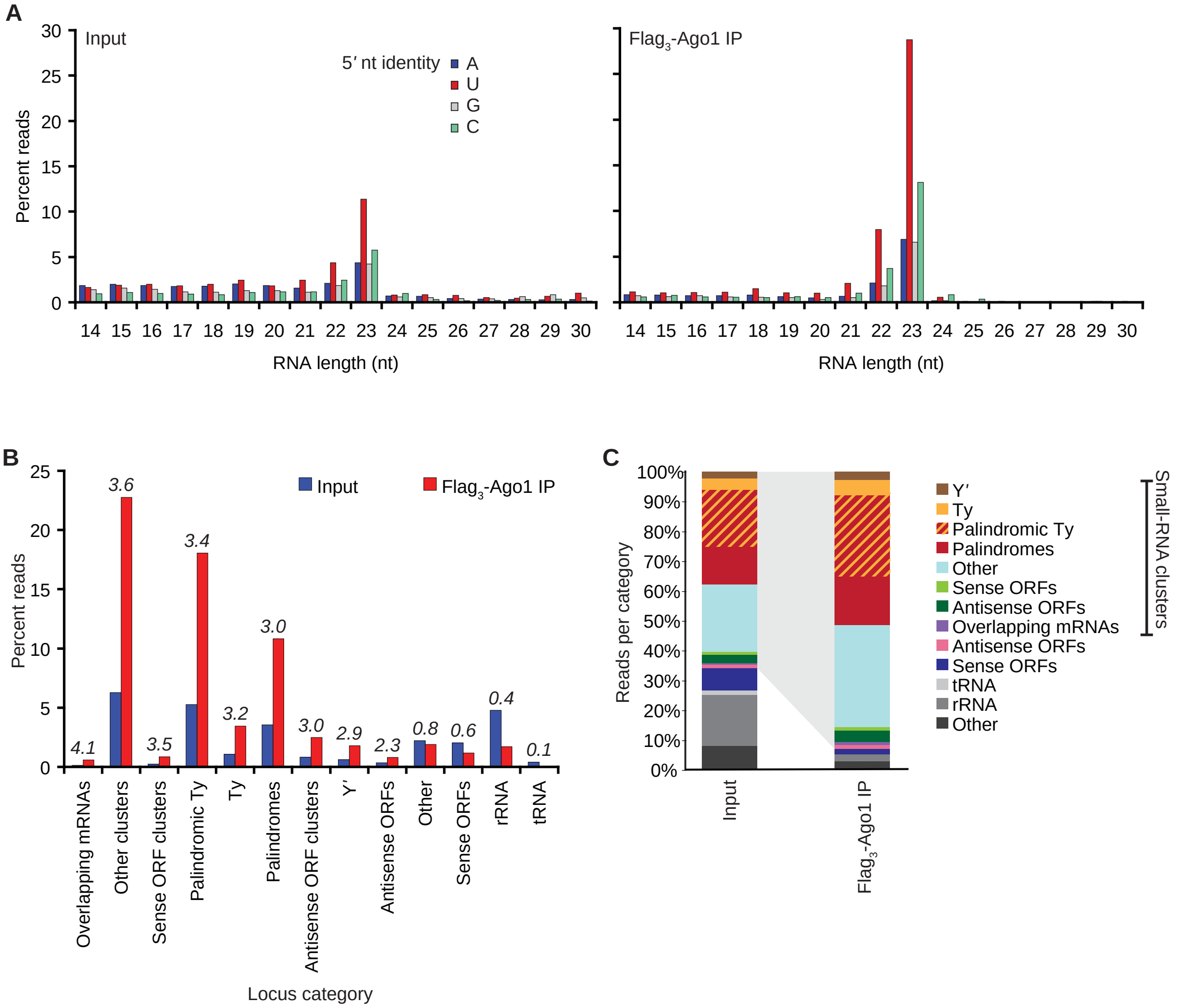


Figure S3. Sequencing of Ago1-associated small RNAs. **(A)** Length distribution of genome-matching sequencing reads representing small RNAs with the indicated 5′ nucleotide. Reads matching rRNA and tRNA are excluded. **(B)** Enrichment analysis of 22–23-nt reads based on genome annotation and alignments of their mapped loci. Italicized numbers above bars represent fold-enrichment calculated as (% of total reads in IP)/(% of total reads in Input). **(C)** Classification of 22–23-nt reads based on genome annotation and alignments of their mapped loci, considering those that map to clusters in a pattern suggestive of siRNAs separately from those that do not. Gray shading indicates the fraction of small RNAs considered to be siRNAs.

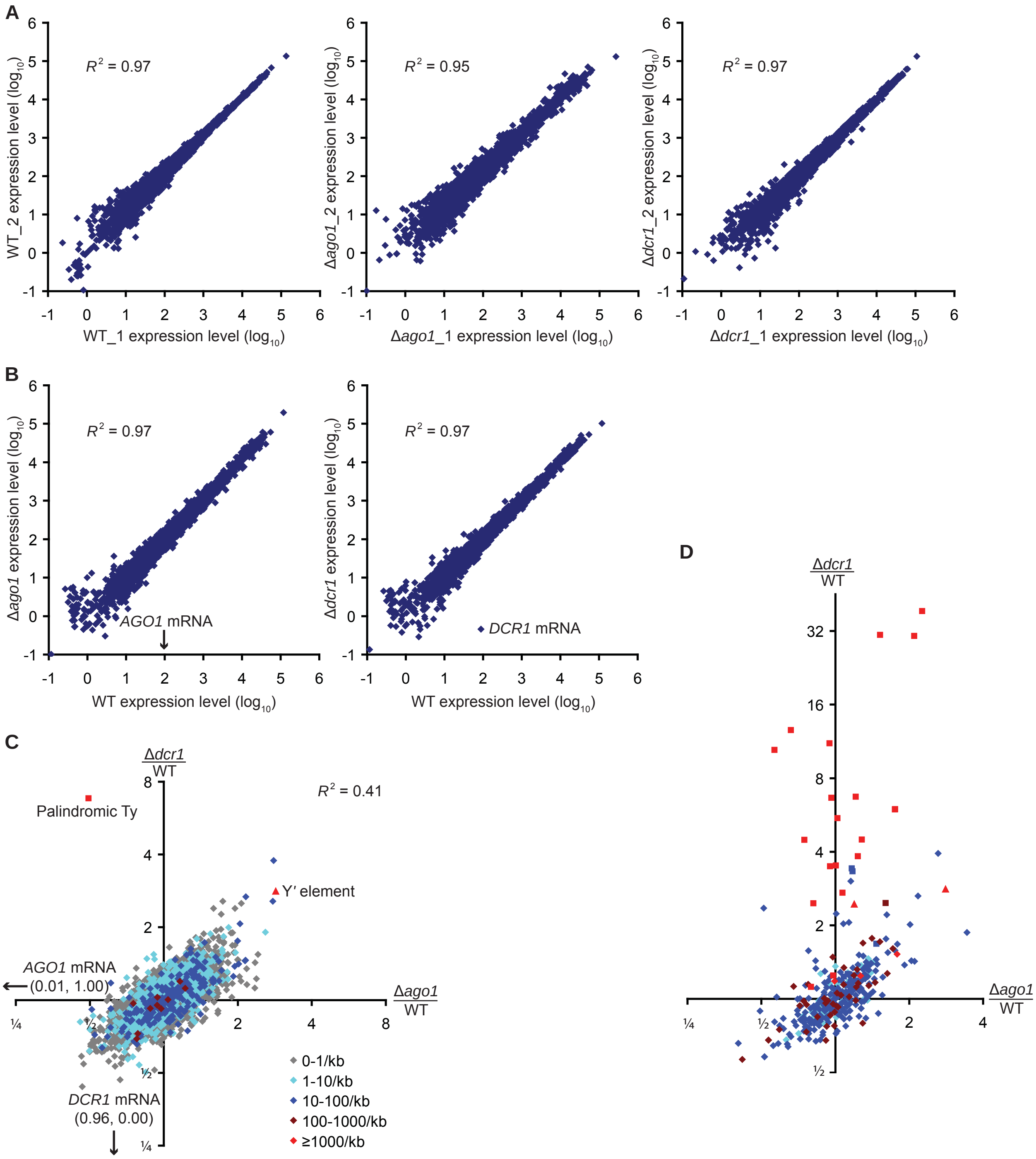


Figure S4. mRNA-Seq analysis of wild-type and RNAi-mutant strains. **(A)** Correlation in transcript abundance between biological replicates. The number of tags mapping to the 5′ half of each annotated ORF was used to estimate the abundance of full-length transcripts. Expression level was calculated as tags per kilobase of coding exon. **(B)** Correlation in transcript abundance between wild-type and RNAi-mutant strains. Plots are as in (A). *AGO1* mRNA had 96.77 tags/kb and 0 tags/kb in WT and Δ*ago1* strains, respectively. **(C)** Plot is as in Figure 3A, except that transcript abundance was calculated using tags across the entire ORF. **(D)** Plot is as in Figure 3C, except that transcript abundance was calculated using tags across the entire transcript.

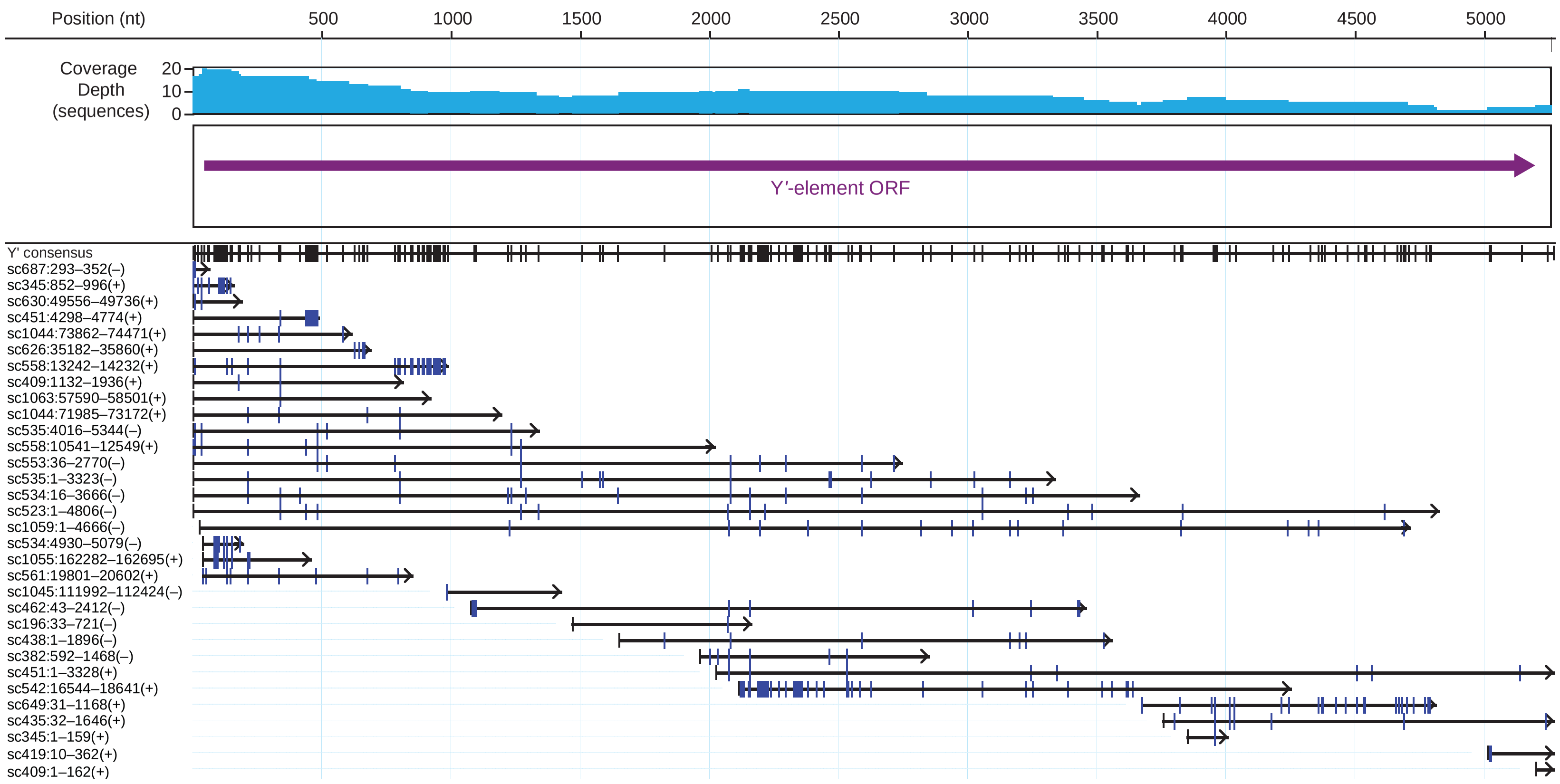


Figure S5. Assembly of a *S. castellii* Y′-element consensus sequence. Y′-element fragments were assembled into a single consensus sequence as described in Materials and Methods. Vertical bars represent single-nucleotide polymorphisms with respect to the majority sequence, many of which fell at the ends of contigs and are presumed to include sequencing errors.

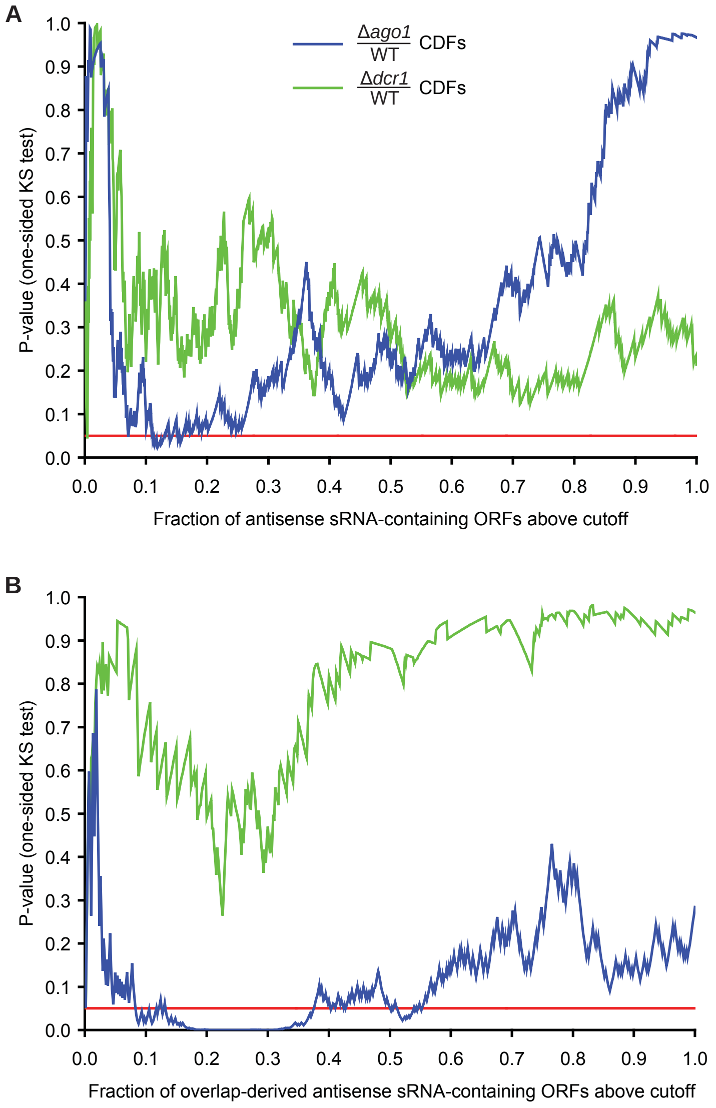


Figure S6. Impact of siRNAs on ORF-containing transcripts. **(A)** Statistical analysis of the impact of small RNAs (sRNAs) mapping antisense to annotated ORFs. ORFs were sorted descending by antisense sRNAs per kb and the significance of transcript down regulation for the ORFs with greater numbers of small RNAs was calculated for the full range of cutoff values. A one-sided KS test was used to compare the distribution of Δ*ago1*/WT (blue) or Δ*dcr1*/WT (green) transcript ratios for ORFs above and below each cutoff. Plotted are the resulting P-values as a function of the cutoff (expressed as the fraction of all antisense-sRNA-containing ORFs included above the cutoff). The red line indicates the $P = 0.05$ significance cutoff. **(B)** Statistical analysis of the impact of sRNAs generated by overlapping convergent gene pairs. ORFs were sorted descending by overlapping-transcript-derived antisense sRNAs/kb and analyzed as in (A).

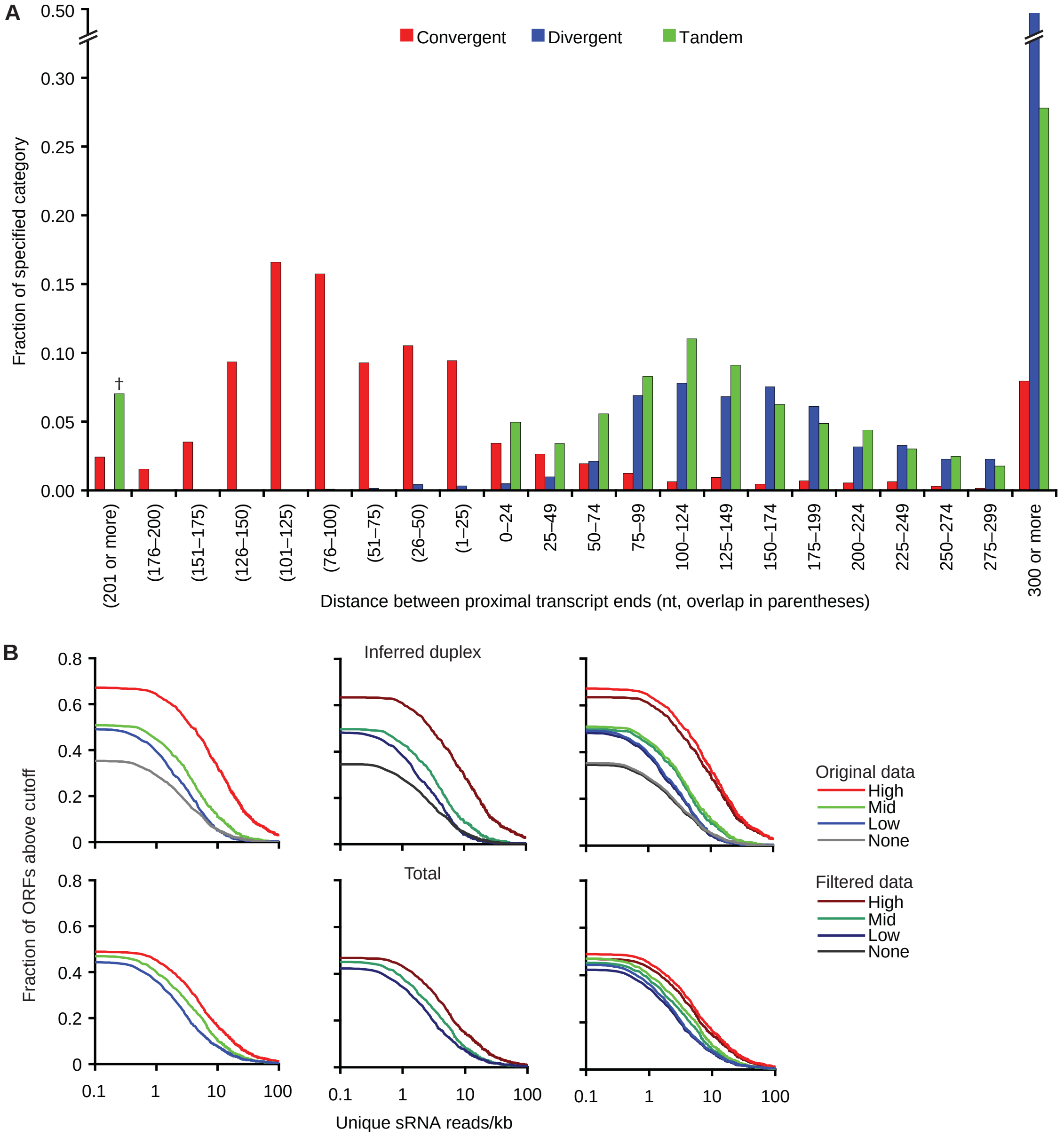


Figure S7. Gene-pair organization and overlap in *S. castellii*. **(A)** Distribution of gene-pair inter-transcript distances. Gene pairs were binned by the distance between 3′-ends (convergent), 5′-ends (divergent), or 3′-end of the upstream gene and 5′-end of the downstream gene (tandem). Plotted is the fraction of gene pairs of a given orientation category that fall within each bin. † For overlapping tandem gene pairs, transcript ends for both genes represent the 5′ and 3′ ends of the contiguous signal observed by mRNA-Seq. Therefore, tandem gene pairs are depicted as overlapping across their length.
(B) Correlation between transcript abundance and small RNA density for annotated ORFs. ORFs were binned according to inferred duplex abundance (estimated as the abundance of the limiting strand; top) or total transcript abundance (sum of sense and antisense tags; bottom). Plotted is the fraction of ORFs within a given bin that have at least as many uniquely matching small RNA reads (on either strand) as the x-axis value. As expected if siRNAs in coding sequences derived from dsRNA precursors formed by sense-antisense transcript pairs, the abundance of ORF siRNAs correlated with the abundance of the inferred duplex. Filtered data excludes all convergent overlapping gene pairs that give rise to small RNAs in the overlap region .

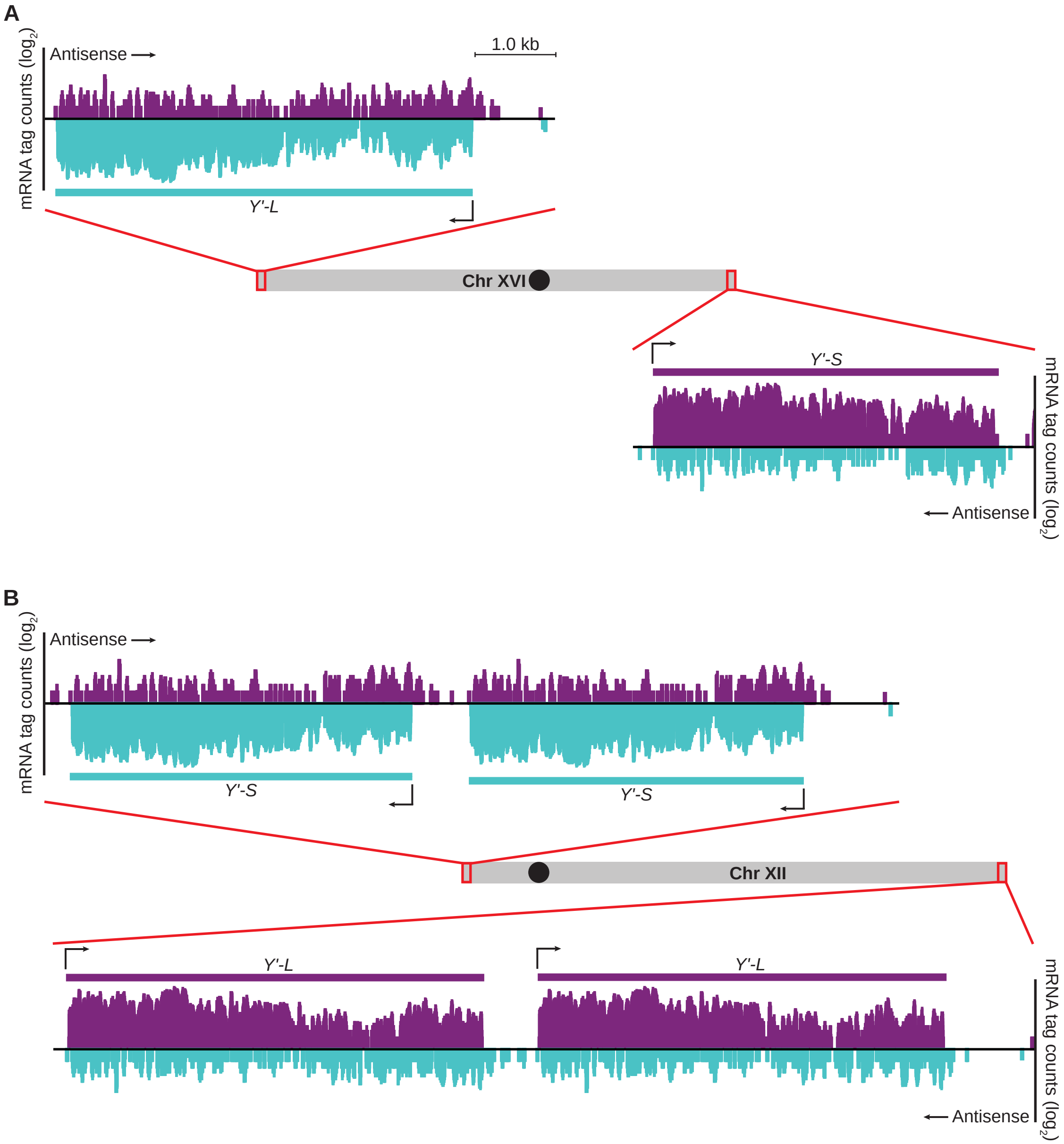


Figure S8. mRNA-Seq analysis of the *S. cerevisiae* Y′ elements. **(A)** Transcripts mapping to chromosome XVI subtelomeres. mRNA-Seq tags were mapped to the reference genome. Tags mapping to the subtelomeric regions of chromosome XVI are shown, with tags contributing to the counts along their entire length. Positions of the vertical axes correspond to the ends of the chromosome. *Y′-L* and *Y′-S* represent the inferred genes corresponding to the long and short isoforms of *S. cerevisiae* Y′ elements, respectively. In *S. cerevisiae*, the telomeres are transcriptionally silenced by Sir2-dependent heterochromatin but still give rise to low levels of cryptic transcripts that are rapidly degraded by the TRAMP and exosome complexes (*S28*). The previously characterized *S. cerevisiae* cryptic telomeric transcripts are ~6.5 kb in length, and begin near chromosome ends and run antisense through the entire Y′-element ORF. The antisense reads we detected across *S. cerevisiae* subtelomeric regions may represent these previously identified cryptic transcripts. **(B)** Transcripts mapping to chromosome XII subtelomeres. Plots are as in (A).

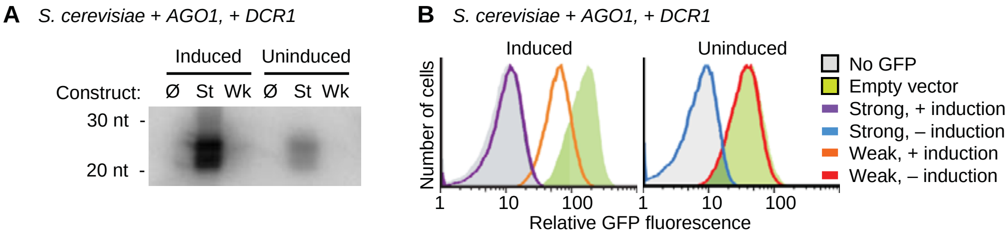


Figure S9. Reconstituting RNAi in *S. cerevisiae*. **(A)** Northern blot for siRNAs antisense to GFP in a *S. cerevisiae* strain expressing *S. castellii AGO1*, *DCR1*, and either no silencing construct (Ø), an integrated strong silencing construct (St), or an integrated weak silencing construct (Wk). Cells were induced in SC media with galactose and raffinose or uninduced in SC media with glucose. **(B)** FACS histograms of GFP fluorescence in *S. cerevisiae* expressing *S. castellii AGO1* and *DCR1* and the indicated silencing constructs. The same cultures were used here for sorting as for RNA collection in (A). In principle the siRNAs and silencing observed under uninduced conditions could be due to leaky expression from the *GAL1* promoter, but these effects are probably attributable to constitutive antisense transcription from a downstream promoter.

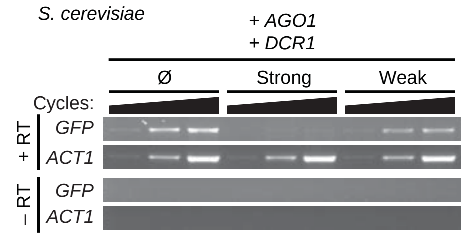
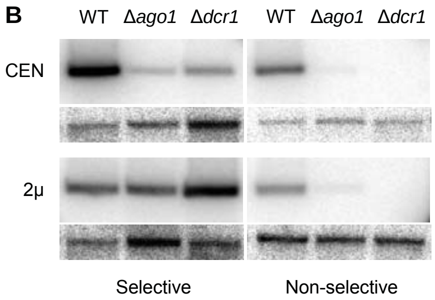


Figure S10. Analysis of *GFP* mRNA in reconstituted RNAi in *S. cerevisiae*. Aliquots from RT-PCR reactions were removed after increasing numbers of PCR cycles (*GFP*: 28, 32, 36; *ACT1*: 24, 28, 32) and visualized by ethidium bromide staining.

A

Plasmid	WT	Δ*ago1*	Δ*dcr1*
CEN	300	5	10
2µ	105	11	7
2µ Ago1	225	7	25
2µ Ago1, induced	169	96	4
2µ Dcr1	169	9	11
2µ Dcr1, induced	142	2	127



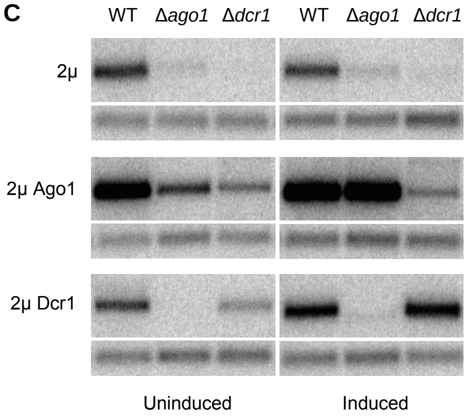
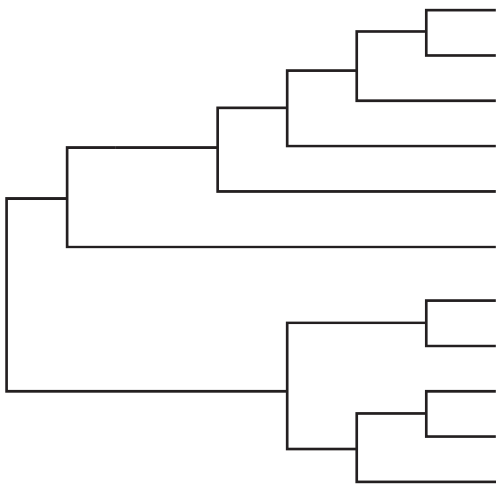


Figure S11. Plasmid instability in RNAi mutants. **(A)** Number of colonies obtained upon transformation of each strain with the plasmid indicated, sum of three independent transformations (table S6). The CEN plasmid was pRS316; 2µ was a 2-micron origin plasmid; 2µ Ago1 and 2µ Dcr1 were 2-micron plasmids expressing Ago1 or Dcr1, respectively, under the *S. cerevisiae* GAL1 promoter. **(B)** Southern blot for abundance of the indicated plasmid in each of the indicated strains. Plasmids (CEN, 2µ) were detected with a probe against the ampicillin-resistance gene; loading controls (thin panels) were probed for a genomic locus. DNA was isolated from cells grown in SC–ura (selective) or YPD (non-selective). **(C)** Southern blot probed as in (B) monitoring rescue of plasmid maintenance phenotype using DNA isolated from cells grown in YPD (uninduced) or YP-galactose (induced).

Δ*ago1* and Δ*dcr1* mutants yielded fewer colonies upon plasmid transformation than did wild-type *S. castellii* (fig. S11A, top two rows). This effect was observed for CEN plasmids (which contained an *S. cerevisiae* centromere sequence and an *S. cerevisiae* chromosomal origin of replication) as well as 2-micron plasmids (which contained the origin of the *S. cerevisiae* endogenous 2-micron circle but no centromere sequence). To distinguish whether this effect reflected a defect in plasmid transformation (plasmid entering the cell) or plasmid maintenance (propagation of the plasmid after entering the cell), we attempted to rescue the defect by transforming wild-type, Δ*ago1*, and Δ*dcr1* strains with plasmids expressing either Ago1 or Dcr1 from an inducible promoter. If the mutant strains were defective in transformation, then these Ago1 and Dcr1 expression plasmids would not enter the cell and thus could not rescue the mutant phenotype. Alternatively, if the mutant strains were defective in plasmid maintenance, then these plasmids would enter the cell, and expression of plasmid-borne Ago1 or Dcr1 in the cognate mutant could rescue maintenance of the expression plasmid itself. When the Δ*ago1* mutant was transformed with the Ago1-expression plasmid and the cells were plated on inducing media, wild-type numbers of colonies were obtained. The same was observed for the Δ*dcr1* mutant transformed with the Dcr1-expression plasmid. This rescue was not observed with the non-cognate plasmids or when expression was not induced (fig. S11A), thereby demonstrating the specificity of the rescue. These results show that RNAi is required for maintenance of *S. cerevisiae* plasmids in *S. castellii*.

We then used Southern blots to monitor plasmid levels. For the CEN plasmid, Δ*ago1* and Δ*dcr1* mutants carried, on average, fewer plasmids per cell relative to wild-type cells, even when grown in selective media (fig. S11B, top). For the 2-micron plasmid, Δ*ago1* and Δ*dcr1* mutants maintained the plasmid at wild-type abundance in selective media, although growth was considerably slower. When allowed to lose plasmid by growth in rich, non-selective media, the mutants lost more plasmid than the wild-type cells did (fig. S11B, bottom). Consistent with the rescue observed when counting colonies (fig. S11A), expressing the relevant protein from the plasmid being monitored rescued the plasmid-maintenance phenotype (fig. S11C). Partial rescue was observed without induction due to leaky expression, but full rescue required induction.



		gypsy family	
	copia family*	Ty3-like	Tca3-like
Saccharomyces cerevisiae	53	2	0
Saccharomyces bayanus	10	0	0
Candida glabrata	0	1	0
Saccharomyces castellii	8	10	8
Kluyveromyces polysporus	18	8	16
Zygosaccharomyces rouxii	0	2	0
Kluyveromyces lactis	2	0	0
Ashbya gossypii	0	1	0
Kluyveromyces waltii	25	3	0
Kluyveromyces thermotolerans	2	0	0
Saccharomyces kluyveri	9	0	0

* Ty1, Ty2, Ty4, Ty5

Figure S12. Approximate copy numbers of retroelements in budding yeast species. Copy numbers were estimated by TBLASTN searches using the Gag-Pol polyprotein as a search query. Intact genes and pseudogenes were counted, but not solo LTRs. *S. castellii* and *K. polysporus* have many more Ty3/gypsy elements (18 and 24 elements, respectively) than those budding yeast species that have lost the RNAi pathway (0–3 elements). Most notably, a subfamily of gypsy elements more similar to *C. albicans* Tca3 (*S29*) than to *S. cerevisiae* Ty3 is found exclusively in species that have retained the RNAi pathway: *S. castellii* and *K. polysporus*, as well as several *Candida* species. The two gypsy subfamilies have been proposed to have different mechanisms for priming minus-strand RNA synthesis (*30*). As in *C. albicans*, many of the members of the gypsy families in *S. castellii* and *K. polysporus* appear to be structurally rearranged. It is possible that selection has favored the retention of these structures as templates for defensive siRNA production.

Table S1. Analysis of small-RNA libraries. Read counts were normalized to the number of genomic matches and separated into different cateories based on genome annotations and alignments. Numbers in parenthesis are percent of reads compared to number of genome matching reads of either all sequence reads or only reads of 21–23-mers.

	S. castellii				*K. polysporus*		*C. albicans*		*S. cerevisiae*	
	wild-type		Δ*ago1*	Δ*dcr1*						
Annotation	all reads	21–23-mers	all reads	all reads	all reads	21–23-mers	all reads	21–23-mers	all reads	21–23-mers
Ty retrotransposons										
Palindromic	27984.0 (12.1)	24419.2 (22.5)	81.0 (0.1)	11.3 (<0.1)	5891.5 (0.7)	5020.2 (1.6)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Non-palindromic	18949.8 (8.2)	16223.0 (14.9)	271.1 (0.3)	9662.6 (0.8)	98666.5 (11.2)	79859.1 (25.5)	881.5 (0.1)	677.0 (0.3)	52.8 (<0.1)	3.9 (<0.1)
Y′ elements	4942.9 (2.1)	4172.5 (3.8)	9.0 (0.0)	50.0 (<0.1)	67354.8 (7.6)	56305.0 (18.0)	–	–	10.8 (<0.1)	1.0 (<0.1)
Zorro3	–	–	–	–	–	–	4809.5 (0.4)	3416.0 (1.6)	–	–
CTA2 family	–	–	–	–	–	–	8991.9 (0.8)	6848.9 (3.1)	–	–
LPF family	–	–	–	–	–	–	2198.5 (0.2)	1604.0 (0.7)	–	–
Other transposon homology *	–	–	–	–	–	–	1719.1 (0.1)	1226.0 (0.6)	–	–
Other palindromes	18322.2 (7.9)	16356.7 (15.0)	51.0 (0.1)	23.5 (<0.1)	20503.5 (2.3)	17039.3 (5.4)	–	–	–	–
ORF clusters	1860.2 (0.8)	1583.9 (1.5)	8.7 (<0.1)	45.3 (<0.1)	11114.0 (1.3)	9170.6 (2.9)	2972.6 (0.25)	2080.1 (1.0)	–	–
Other siRNA clusters	15191.2 (6.5)	13267.8 (12.2)	87.0 (<0.1)	6977.3 (0.6)	39390.4 (4.5)	31612.7 (10.1)	8487.0 (0.7)	5367.5 (2.5)	–	–
ORFs	2829.4 (1.2)	1971.5 (1.8)	123.7 (0.1)	23616.4 (1.9)	9094.7 (1.0)	5568.7 (1.8)	19253.8 (1.6)	8131.6 (3.7)	3783.1 (1.0)	700.2 (1.0)
Pseudogenes	n.d.	n.d.	n.d.	n.d.	73.4 (<0.1)	48.0 (<0.1)	n.d.	n.d.	5.2 (<0.1)	2.0 (<0.1)
Centromeres	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	15.0 (<0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Telomeres	n.d.	n.d.	n.d.	n.d.	2284.2 (0.3)	1817.5 (0.6)	n.d.	n.d.	0.2 (<0.1)	0.0 (0.0)
Non-protein coding RNAs										
rRNA	103905.9 (44.7)	19039.8 (17.5)	74687.1 (76.2)	746407.5 (62.7)	525501.2 (59.6)	80098.5 (25.6)	990056.6 (83.2)	148932.0 (68.0)	310004.5 (84.9)	63278.7 (87.4)
tRNA	3550.4 (1.5)	565.2 (0.5)	2635.4 (2.7)	98132.5 (8.2)	63408.8 (7.2)	8267.0 (2.6)	39074.3 (3.3)	6225.0 (2.8)	38398.3 (10.5)	6140.0 (8.5)
Other ncRNA	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	30297.8 (2.5)	6309.0 (2.9)	2339.4 (0.6)	431.0 (0.6)
Other	34680.1 (14.9)	11169.5 (10.3)	19997.0 (20.4)	305659.6 (25.7)	30565.8 (3.5)	13249.4 (4.2)	80747.5 (6.8)	28135.0 (12.8)	10691.7 (2.9)	1869.2 (2.6)
Total genome-matching reads	232226.0 (100)	108769.0 (100)	97951.0 (100)	1190586.0 (100)	881041.0 (100)	312884.0 (100)	1189505.0 (100)	218952.0 (100)	365286.0 (100)	72426.0 (100)
Total reads	510234		528020	4265687	1777358		3334405		1227275	

* Hypothetical protein CaO19.7545, potential SWIM zinc finger protein very similar to CaP19.4241, C terminus is same as Rel-associated hypothetical protein, MULE transposase domain;
Hypothetical protein CaO19.6608, weak similarity to C. albicans Cirt2 transposase

n.d. - not determined

– - not applicable

Table S2. Analysis of small-RNA libraries from input and Flag$_3$-Ago1 IP datasets. Read counts were normalized to the number of genomic matches and separated into different cateories based on genome annotations and alignments. Numbers in parentheses are percent of reads compared to number of genome matching reads of either all sequence reads or only reads of 22–23-mers.

	Input				Flag$_3$-Ago1 IP			
	all reads		22–23-mers		all reads		22–23-mers	
Ty retrotransposons								
Non-palindromic Ty	23892.9	(1.7)	15310.0	(3.8)	53793.7	(4.5)	40818.3	(5.2)
Palindromic Ty	110728.9	(7.7)	75986.0	(19.0)	281476.6	(23.7)	213975.9	(27.2)
Y′ elements	14058.9	(1.0)	9037.7	(2.3)	29293.5	(2.5)	21335.9	(2.7)
Other palindromes	70149.7	(4.9)	51205.4	(12.8)	164269.9	(13.9)	128153.5	(16.3)
ORF clusters								
Sense ORF clusters	7172.8	(0.5)	3530.2	(0.9)	13191.2	(1.1)	10197.4	(1.3)
Antisense ORF clusters	17705.9	(1.2)	11774.5	(2.9)	38322.0	(3.2)	29345.9	(3.7)
Overlapping mRNAs	4958.4	(0.3)	2106.5	(0.5)	9251.8	(0.8)	7028.0	(0.9)
Other siRNA clusters	135377.1	(9.4)	90559.8	(22.6)	368810.2	(31.1)	269707.6	(34.3)
Open reading frames								
Sense ORFs	244127.2	(16.9)	29220.9	(7.3)	49693.8	(4.2)	13907.4	(1.8)
Antisense ORFs	20280.8	(1.4)	5033.7	(1.3)	18576.6	(1.6)	9380.3	(1.2)
tRNA	66699.1	(4.6)	5629.7	(1.4)	3573.7	(0.3)	244.4	(<0.1)
rRNA	479478.7	(33.3)	68866.7	(17.2)	101753.5	(8.6)	20168.2	(2.6)
Other	245917.5	(17.1)	32093.0	(8.0)	53441.5	(4.5)	22434.4	(2.9)
Total genome-matching reads	1440548	(100)	400354	(100)	1185448	(100)	786697	(100)
Total reads	4310251				4102562			

Table S3. siRNA-producing loci and their transcripts. Start and end represent the inferred 5′ and 3′ ends of the transcript, respectively. Coverage is the percentage of nucleotides in the transcript (excluding nucleotides added during fine mapping) that are represented by mRNA-Seq tags.

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc1:429–661	+		429	661	82.4
	–		661	429	85.8
sc2:68–238	+		58	238	25.4
	–	*NCS33*	468	58	96.8
sc196:342–609	+		342	609	0.0
	–		649	122	95.3
sc276:74–1288	+		34	1303	33.8
	–	*NCS42*	1288	24	83.2
sc335:46–172	+		46	172	48.8
	–		172	46	46.5
sc335:463–721	+		463	721	9.7
	–	*NCS84*	873	363	98.2
sc345:1–432	+		1	441	78.5
	–		432	1	48.4
sc360:1541–1647	+	*NCS85*	1351	1672	100.0
	–		1647	1541	0.0
sc365:495–600	+		44	620	98.9
	–		870	475	86.6
sc375:37–696	+		37	766	94.1
	–		696	37	26.5
sc377:1723–2191	+		1723	2191	16.0
	–		2301	1343	95.9
sc382:317–1193	+		317	1193	57.8
	–		1193	244	78.9
sc388:360–581	+		360	581	22.5
	–		601	250	86.9
sc388:833–1511	+		833	1511	14.7
	–		1511	833	7.4
sc391:1088–1185	+		1078	1235	90.5
	–		1235	1068	81.5
sc391:1935–2433	+		1935	2433	51.5
	–		2453	1895	72.8
sc409:577–712	+		557	722	84.3
	–		712	577	18.4
sc412:1789–1960	+		1643	2138	100.0
	–		1960	1789	0.0
sc413:1937–2027	+		1157	2037	99.9
	–		2717	1899	100.0
sc419:1286–1358	+	*NCS72*	1262	1518	100.0
	–		1398	1286	44.2
sc425:395–1034	+		395	1034	15.9
	–	*NCS21*	1184	225	93.2
sc425:1295–2077	+		1295	2077	10.5
	–	*NCS19*	2127	1215	86.6
sc435:32–1092	+		32	1552	90.9
	–		1092	32	73.6
sc435:1977–2192	+		1917	2312	83.6
	–		2192	1977	23.1
sc435:2434–2662	+		2434	2772	83.2
	–		2662	2434	21.8
sc436:1030–1361	+		990	1881	100.0
	–		1361	1030	0.0
sc437:3822–4066	+		3815	4072	98.8
	–	*NCS77*	4396	3822	96.9
sc438:1–1441	+		1	1441	76.3
	–		1454	1	73.5
sc441:660–812	+		660	812	0.0
	–	*NCS39*	895	450	96.4
sc451:79–2774	+		66	3234	84.7
	–		2774	79	77.1
sc455:5844–5967	+		5844	5967	0.0
	–	*NCS61*	6309	5442	99.0
sc462:65–1252	+		65	1252	85.9
	–		1252	65	76.4
sc462:1642–2097	+		1642	2097	0.0
	–		2098	1362	91.3
sc471:5550–5778	+		5550	5778	28.4
	–	*NCS76*	5870	5190	99.9
sc471:6204–6659	+	*NCS53*	6204	6734	87.1
	–		6659	5114	86.1
sc473:1270–2094	+		1055	2134	100.0
	–		2094	1270	11.6
sc473:2266–2389	+		666	2779	99.7
	–		2389	2266	0.0
sc474:2097–2411	+		2010	2960	100.0
	–		2411	2097	31.7
sc474:5702–6100	+		5702	6140	45.8
	–		6100	5177	96.8
sc476:6759–7299	+		6679	7389	86.2
	–		7299	6759	16.3
sc476:7961–8182	+		7941	8292	83.2
	–		8182	7961	22.5
sc485:8485–9090	+		8475	9180	86.3
	–		9090	8485	23.9
sc490:8033–8385	+		8033	8385	7.1
	–		8410	7163	100.0
sc490:9427–9629	+	*NCS74*	9127	9957	100.0
	–		9629	9427	12.3
sc493:8255–8723	+		8255	8743	32.7
	–		8893	7675	95.7
sc497:9649–10263	+		9649	10263	68.5
	–	*NCS10*	10263	9189	98.3
sc502:4–583	+		1	583	21.4
	–		623	1	100.0
sc502:753–1055	+		753	1055	8.3
	–		1075	625	100.0
sc506:13783–13836	+		13783	14092	94.8
	–		13837	13783	100.0
sc508:7282–7527	+		7276	7533	100.0
	–	*NCS79*	7760	7282	97.9
sc509:188–473	+		158	503	70.2
	–		533	188	54.9
sc509:9632–9853	+		9632	9963	77.4
	–		9853	9632	11.3
sc509:10321–10542	+		10301	10602	87.4
	–		10542	10321	22.5
sc511:11–213	+		11	213	0.0
	–		213	1	46.5
sc517:Feb–16	+		1	3786	73.6
	–	*NCS35*	2516	1	71.7
sc517:4869–5249	+		4779	5279	49.7
	–		5249	4859	72.4
sc517:5592–6682	+		5572	7096	97.8
	–		6682	5592	54.5
sc517:6922–7017	+		6144	7040	99.8
	–		7017	6922	26.0
sc517:7607–7752	+		7607	7752	0.0
	–		7752	7327	83.8
sc519:294–618	+		294	618	21.8
	–		1118	290	98.2
sc522:1143–1744	+		1143	1744	8.3
	–		1744	1083	75.4
sc523:Oct–05	+		1	2705	76.3
	–		2718	1	82.2
sc523:2962–3417	+		2962	3417	0.0
	–		3418	2742	92.2
sc523:5224–5445	+		5224	5445	11.3
	–		5495	5151	95.0
sc525:4218–4572	+		4218	4572	7.6
	–		4945	3958	100.0
sc527:4128–4728	+		4128	4728	50.7
	–	*NCS17*	4768	4028	99.2
sc528:14255–14336	+		13315	14342	100.0
	–		14676	14245	97.7
sc534:69–1432	+		68	1432	87.5
	–		1432	1	76.7
sc534:1822–2277	+		1822	2277	0.0
	–		2278	1592	89.5
sc534:8952–9353	+		8912	9353	17.8
	–	*NCS67*	9503	8922	80.4
sc534:9993–10089	+		9993	10089	0.0
	–		10089	9863	84.6
sc534:10195–10626	+		10195	10626	0.0
	–		10626	10115	71.1
sc535:1–1222	+		1	1222	71.5
	–		1235	1	76.3
sc535:1479–1920	+		1479	1920	0.0
	–		1940	1249	86.7
sc537:2001–2167	+		1906	3987	99.9
	–		2167	2001	0.0
sc539:159–405	+	*NCS80*	1	405	97.5
	–		411	153	100.0
sc539:1276–1703	+		1276	1703	0.0
	–		1825	1255	96.8
sc542:764–929	+		764	929	0.0
	–		929	764	51.2
sc542:1579–1684	+		1579	1684	0.0
	–		1684	1579	34.9
sc542:6091–6205	+		5091	6221	100.0
	–		6205	6081	63.2
sc542:14947–15985	+		14947	16130	91.3
	–		15985	14947	24.5
sc542:16197–18913	+		16197	19025	63.4
	–		18913	16156	89.2
sc547:13848–14000	+		13848	14000	38.6
	–		14000	13828	50.3
sc549:16769–17410	+		16569	17440	80.3
	–		18397	16769	83.2
sc553:36–669	+		36	669	53.0
	–		1231	36	79.8
sc553:926–1381	+		926	1381	0.0
	–		1382	696	92.9
sc553:3188–3409	+		3188	3409	22.5
	–		3429	3078	86.9
sc558:9896–10121	+		9896	10191	89.2
	–		10121	9896	11.1
sc558:11930–12385	+		11920	12535	90.1
	–		12385	11930	0.0
sc558:12604–12822	+		12604	12932	78.1
	–		12822	12604	11.4
sc559:825–1331	+		813	1481	100.0
	–		1331	815	16.4
sc559:8888–9015	+		8878	9055	55.6
	–		9255	8888	73.1
sc561:20642–20871	+		20602	20947	93.4
	–		20871	20642	10.9
sc562:19159–19222	+		18169	19262	99.8
	–		19562	19149	95.4
sc565:20405–20997	+		20405	21027	36.9
	–	*NCS51*	20997	20365	77.4
sc565:21201–21383	+		21181	21383	82.8
	–		21413	21191	87.4

Continued on next page

Table S3, continued

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc565:21580–21803	+		21520	21891	66.1
	–		21803	21510	85.0
sc567:20621–20981	+		20141	20981	80.5
	–		21231	20621	97.2
sc568:18314–18566	+		18283	18616	95.5
	–		18566	18184	65.0
sc575:5776–5935	+		4986	5947	100.0
	–		6075	5776	88.0
sc578:13092–13331	+		13092	13331	45.8
	–		13363	12852	95.7
sc580:2233–4179	+	*NCS31*	2233	4179	98.9
	–		4179	2213	26.0
sc582:23773–24238	+	*NCS43*	23733	24309	78.3
	–		24638	23662	99.5
sc582:24415–24595	+		24365	24645	76.9
	–		24638	23662	99.5
sc587:198–326	+		194	326	96.9
	–	*NCS27*	506	198	90.3
sc587:663–788	+		663	788	84.1
	–	*NCS28*	833	543	98.6
sc588:7386–7753	+		6836	7843	99.6
	–		8453	7306	94.2
sc588:30087–30218	+		30087	30218	0.0
	–		30348	29567	100.0
sc589:3374–3461	+		2794	3463	100.0
	–		5031	3354	100.0
sc589:20446–20656	+		20446	20656	0.0
	–		20826	20336	92.3
sc596:1304–1461	+		1304	1461	15.8
	–		1501	1064	100.0
sc600:9619–10607	+		9610	10607	1.6
	–	*NCS13*	11697	9287	99.9
sc600:10849–11141	+		10849	11141	0.0
	–	*NCS16*	11695	9779	100.0
sc604:4257–4935	+		4246	6113	99.0
	–		4935	4237	51.9
sc608:25105–25468	+		25095	25468	81.8
	–	*NCS81*	25650	25105	97.3
sc610:2400–2460	+		1999	2483	100.0
	–		2470	2400	62.3
sc610:21566–21806	+		21566	21846	17.8
	–		22616	21512	100.0
sc610:22108–22644	+		22068	22664	44.7
	–		22654	21512	99.0
sc618:9475–9557	+		7925	9574	100.0
	–		10191	9445	98.8
sc621:39720–39823	+		39720	39823	0.0
	–		39912	39090	99.8
sc621:40611–40703	+		40611	40703	0.0
	–		40799	39931	98.6
sc622:33321–33622	+		33201	33622	67.5
	–	*NCS69*	33882	33181	94.2
sc622:34331–34482	+	*NCS65*	34041	34703	97.7
	–		34482	34311	29.1
sc622:34908–35226	+	*NCS59*	34388	35226	95.5
	–		35246	34908	36.9
sc622:38796–39212	+		38796	39332	75.6
	–		39212	38796	24.0
sc622:39454–39682	+		39454	39742	80.3
	–		39682	39454	21.8
sc622:39899–40113	+		39889	40223	84.2
	–		40113	39899	23.3
sc623:4134–4482	+		4134	4482	7.2
	–		4543	3900	98.0
sc623:4774–5347	+		4774	6473	96.7
	–		5347	4764	35.9
sc623:33964–34329	+		32584	34329	99.4
	–		34334	33964	82.8
sc625:38090–38215	+		37338	38277	100.0
	–		38275	38090	55.4
sc626:1404–1897	+		1404	1907	28.8
	–		1897	1384	59.1
sc626:4673–7708	+		4623	7859	81.6
	–		7718	4653	26.3
sc626:9067–9601	+	*NCS58*	8967	9611	66.4
	–	*NCS57*	9601	8997	83.0
sc626:10146–10513	+		10146	10513	0.0
	–		10513	10146	0.0
sc626:14217–14526	+		14217	14526	27.4
	–		14526	14077	90.9
sc626:34543–34761	+		34523	34771	85.1
	–		34761	34543	22.8
sc630:49045–49132	+		48985	49144	94.9
	–		49132	49045	28.4
sc634:22393–22468	+		21616	22514	100.0
	–		23734	22391	99.3
sc635:14611–14845	+		14605	14858	100.0
	–		17485	14611	99.7
sc635:22395–24397	+		22395	24417	26.1
	–		24460	22275	92.9
sc639:6048–6314	+		5998	6604	100.0
	–		6314	6048	18.7
sc639:6526–6571	+		5996	6611	100.0
	–		6571	6516	44.6
sc639:11902–12807	+		11902	12847	74.2
	–		12807	11902	5.5
sc639:14024–14099	+		13733	14109	90.4
	–		14284	14001	100.0
sc640:9977–10500	+		9977	10500	0.0
	–		10665	9797	99.7
sc640:41728–41853	+		41238	41863	100.0
	–		41993	41728	94.7
sc641:11318–11523	+		10228	11523	88.7
	–		12033	11278	99.9
sc648:17560–17977	+		17560	17977	6.0
	–	*NCS49*	18172	16246	100.0
sc649:36–1264	+		36	1264	60.6
	–		1294	36	98.8
sc651:4–656	+	*NCS9*	1	1886	99.9
	–		656	1	4.9
sc651:1114–1851	+	*NCS8*	1	1891	99.9
	–	*NCS11*	1851	1114	55.4
sc651:3275–3339	+		2987	3355	100.0
	–		3399	3275	69.6
sc651:20127–20473	+		19223	20503	99.8
	–		20473	20127	0.0
sc653:7042–7296	+		6422	7579	94.6
	–		7296	7032	46.8
sc654:28406–28550	+		28386	28550	55.2
	–		29391	28226	99.8
sc654:55126–55305	+		54366	55305	96.2
	–		56633	55126	99.5
sc658:1329–1442	+		979	1493	96.9
	–		2808	1289	100.0
sc660:25974–26295	+		25574	26295	99.0
	–		26621	25964	99.2
sc660:34132–34683	+		34012	34868	99.8
	–		34784	34102	73.2
sc666:20702–20809	+		20202	20809	100.0
	–		20931	20702	100.0
sc666:32177–33508	+	*NCS23*	32167	33588	72.7
	–	*NCS24*	33508	32177	38.8
sc668:757–936	+		757	1000	82.1
	–		936	757	66.7
sc668:8927–8993	+		8587	9023	97.5
	–		11628	8907	99.9
sc668:14781–14990	+		14781	15020	57.1
	–		15110	14667	98.9
sc668:15296–15545	+		15286	15545	22.7
	–		15668	15186	99.6
sc673:19040–19451	+		19040	19451	45.6
	–		19451	18869	96.4
sc676:55494–55600	+		55454	55630	59.9
	–		55600	55494	23.4
sc676:55766–55886	+		55746	55886	35.5
	–		55886	55766	0.0
sc676:64085–64882	+		63565	64882	60.5
	–	*NCS40*	65004	64085	77.1
sc682:2205–2503	+		1985	3114	94.9
	–		2503	2205	0.0
sc682:24064–24178	+		23038	24198	98.8
	–		25038	24052	100.0
sc682:71530–71636	+		71040	71646	100.0
	–		71656	71530	55.1
sc683:12949–13534	+		12909	13584	58.1
	–	*NCS41*	13545	12949	86.4
sc686:19278–19559	+		19218	19589	78.0
	–		19639	19278	61.6
sc686:23938–26333	+		23918	26363	59.3
	–	*NCS26*	26334	23938	65.2
sc687:39–239	+		39	239	0.0
	–		239	29	53.1
sc687:770–991	+		770	991	22.5
	–		1011	760	84.1
sc687:1246–1450	+		1246	1460	23.3
	–		1490	1186	86.6
sc687:7487–8354	+		7417	8354	19.6
	–	*NCS60*	8704	7356	92.4
sc687:9826–10082	+		9826	10082	10.9
	–		10082	9826	0.0
sc687:59448–59803	+		59428	59833	97.3
	–		60023	59418	99.0
sc687:80365–80851	+		80365	80851	10.3
	–		81831	80293	96.4
sc687:81469–81624	+		81469	81624	0.0
	–		81825	80487	100.0
sc687:83334–83489	+	*NCS70*	83014	83622	98.3
	–		83489	83334	0.0
sc689:59311–59545	+		59311	59545	0.0
	–		59922	59006	100.0
sc695:29108–29282	+		27808	29312	94.0
	–	*NCS47*	29292	29108	96.0
sc696:31624–32479	+		31624	32479	15.8
	–		32479	31564	44.3
sc696:54775–54953	+		54775	54953	14.0
	–		55433	54695	100.0
sc696:58976–59060	+		58346	59079	99.9
	–		60430	58956	99.7
sc696:100414–100848	+	*NCS36*	99638	101123	99.9
	–		100848	100414	19.8
sc696:101156–101809	+	*NCS2*	101136	101809	94.8
	–	*NCS1*	101912	101156	98.4

Continued on next page

Table S3, continued

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc697:26569–26676	+		25879	26686	99.7
	–		27214	26569	98.4
sc701:29781–29871	+		29356	29891	100.0
	–		29941	29781	53.4
sc701:57551–57929	+	*NCS64*	57431	57929	87.5
	–		57939	57551	40.4
sc701:83724–83892	+		82724	83892	93.7
	–		84282	83714	100.0
sc709:24929–25083	+		24738	25083	89.9
	–		25357	24920	100.0
sc709:28052–28465	+		27952	28485	74.0
	–		29028	28002	96.4
sc709:43781–43874	+		43171	43874	92.0
	–		44304	43751	100.0
sc709:69749–70066	+		69159	70396	100.0
	–		70066	69739	4.7
sc709:86133–86705	+		86003	86766	70.2
	–		86825	85513	91.6
sc710:8141–8912	+		7781	9082	93.1
	–		8912	8121	82.1
sc710:12032–12080	+		12002	12080	82.3
	–		12080	12032	0.0
sc710:12346–12436	+		12236	12436	80.1
	–		12436	12346	0.0
sc710:12948–13955	+	*NCS5*	12948	14355	99.4
	–		13955	12948	25.0
sc710:36397–37654	+		36327	37704	86.4
	–		37654	36397	7.9
sc710:37835–38239	+		37813	38259	96.2
	–		38249	37835	21.9
sc710:38873–40259	+		38873	40279	11.1
	–	*NCS34*	40259	38864	98.7
sc710:126921–127123	+		126651	127123	87.9
	–		128033	126921	97.6
sc712:108122–108755	+		108122	108755	27.9
	–		108835	107733	96.0
sc712:125879–125996	+	*NCS55*	125569	126026	100.0
	–		126958	125879	99.0
sc714:111363–111513	+		111353	111513	15.5
	–	*NCS45*	111523	111353	100.0
sc714:112909–112975	+		112049	112988	99.9
	–		113325	112909	90.2
sc716:28464–28791	+	*NCS14*	28444	29886	96.8
	–		28791	28444	14.4
sc716:29088–29402	+	*NCS15*	28657	29886	99.8
	–		29402	29088	0.0
sc716:159394–159521	+		158574	159531	99.4
	–		160538	159389	100.0
sc719:52927–53456	+		52927	53456	0.0
	–		53603	52707	91.9
sc719:80550–80667	+		78360	80677	100.0
	–		81447	80550	95.5
sc719:80983–81279	+		80983	81309	15.3
	–		81449	80582	98.8
sc719:145408–145856	+		145408	145907	67.7
	–		145866	145408	18.3
sc719:188938–189063	+		187376	189093	98.8
	–		190169	188926	99.5
sc720:16982–17198	+		16912	17198	84.7
	–		17848	16962	100.0
sc720:17512–17778	+		17512	17788	70.8
	–		17848	16972	100.0
sc720:86102–86250	+		85842	86250	100.0
	–		87980	86102	99.4
sc721:29938–30311	+		29938	30311	0.0
	–		30351	29888	59.5
sc721:30729–30988	+		30699	30988	64.5
	–		31018	30719	70.0
sc721:31159–32062	+		31149	32122	89.8
	–	*NCS44*	32082	31159	74.8
sc721:42216–43245	+		42201	43340	100.0
	–		43255	42216	25.7
sc721:83771–84102	+		83771	85392	99.3
	–		84102	83771	30.1
sc721:87525–88833	+		87525	88833	10.5
	–		89781	87375	98.0
sc721:104062–104151	+		102375	104181	99.9
	–		105461	104052	99.4
sc721:140124–140427	+		139464	140450	100.0
	–		140577	140124	98.2
sc721:145814–145981	+		145784	145981	69.1
	–		147571	145814	97.1
sc721:191325–191431	+		191135	193881	100.0
	–		191431	191325	0.0
sc721:193779–194025	+		192279	194025	99.4
	–		194030	193773	100.0
sc721:210477–210897	+		210430	211350	100.0
	–		210897	210477	5.9
sc721:277776–277936	+		277616	278030	100.0
	–		277986	277776	39.8
sc1001:29178–29235	+		28748	29235	97.5
	–		29765	29148	100.0
sc1003:4321–4694	+		4260	4704	96.4
	–		5464	4321	97.6
sc1004:15739–15963	+		14289	15963	98.0
	–		16687	15739	91.4
sc1005:16254–16496	+		15514	16536	99.5
	–		16496	16214	74.9
sc1007:9377–9627	+		9370	12730	100.0
	–		9627	9377	0.0
sc1007:20870–21184	+		20830	21254	85.9
	–		21184	20870	46.0
sc1007:47015–47100	+		43965	47137	100.0
	–		47963	47015	99.0
sc1007:82752–82976	+		82752	82976	23.1
	–		82976	82752	0.0
sc1009:8540–8765	+		8235	8765	89.2
	–		9265	8531	100.0
sc1009:34804–35229	+		34804	35499	67.5
	–		35229	34784	10.8
sc1009:48729–49028	+		48729	49058	42.4
	–		49078	48634	99.8
sc1009:49220–49458	+		49210	49471	98.8
	–	*NCS68*	50968	49220	99.5
sc1013:2878–3451	+		2878	3451	4.4
	–		3467	1323	99.4
sc1014:32074–32342	+		32004	32422	74.7
	–		32342	32064	41.9
sc1014:33081–33362	+		32991	33373	100.0
	–		33372	33051	39.1
sc1014:55954–56360	+	*NCS30*	55804	56430	100.0
	–		56360	55954	0.0
sc1014:57720–60475	+		57700	60475	38.1
	–	*NCS25*	60505	57720	60.7
sc1014:62249–63722	+		62249	63722	10.2
	–		63722	62239	11.0
sc1014:69378–71630	+	*NCS4*	69378	71694	76.4
	–	*NCS3*	71630	69318	99.0
sc1014:84431–84529	+		83571	84531	100.0
	–		84875	84412	100.0
sc1014:118810–119191	+		118810	119561	90.2
	–		119211	118800	36.4
sc1017:23149–23389	+		23119	23469	91.7
	–		23389	23149	25.3
sc1017:23803–23901	+		23573	23991	92.6
	–		23943	23763	87.2
sc1018:68–470	+		68	470	6.2
	–		470	58	63.9
sc1018:39398–39759	+		39391	39759	81.8
	–	*NCS82*	39929	39398	98.1
sc1020:3716–4013	+		3716	4013	0.0
	–	*NCS71*	4116	3606	95.1
sc1020:4270–4488	+		4270	4488	52.1
	–		4568	4240	83.9
sc1020:22731–23586	+		22731	23616	71.8
	–		24306	22729	100.0
sc1020:52550–53355	+		52218	53355	90.8
	–		53355	52490	59.0
sc1020:54039–54241	+		53739	55041	100.0
	–		54241	54039	12.3
sc1020:54545–54748	+		53745	55041	100.0
	–		54748	54545	0.0
sc1020:54976–57752	+		54966	59245	72.9
	–	*NCS32*	57752	54956	75.3
sc1020:79835–80260	+		79795	80410	72.6
	–		80260	79835	5.9
sc1020:87293–87522	+		87253	87634	92.9
	–		87522	87293	10.9
sc1022:20893–21016	+		20443	21016	91.5
	–		22516	20881	100.0
sc1023:2454–2826	+		2434	2876	100.0
	–		2876	2454	29.3
sc1024:12371–12696	+		12261	12700	100.0
	–		12706	12341	31.7
sc1025:24448–24480	+		24348	24490	93.0
	–		24480	24448	78.8
sc1027:17084–17242	+		17044	17413	83.2
	–		17242	17084	31.4
sc1028:2027–2262	+		1269	2262	88.0
	–		2872	2027	91.4
sc1028:39441–39939	+		39441	39939	60.7
	–		39959	39188	98.3
sc1029:14930–15299	+	*NCS62*	14930	15459	96.0
	–		15299	14840	56.1
sc1029:26159–26356	+	*NCS54*	25899	26356	83.2
	–		27436	26151	100.0
sc1029:38296–38412	+		37995	38455	99.8
	–		39057	38286	99.3
sc1030:29320–30386	+	*NCS38*	29320	30446	55.2
	–		30386	29260	48.7
sc1031:35404–35604	+		35404	35604	49.8
	–		36286	35301	99.9
sc1033:24770–24913	+		24709	25354	100.0
	–		24963	24770	36.6
sc1033:49898–50136	+	*NCS83*	49888	50146	100.0
	–	*NCS75*	50502	49848	97.3
sc1034:1960–2116	+		1940	2116	75.1
	–		2168	1750	97.4
sc1034:37129–37556	+		37129	37556	0.0
	–		37678	37108	96.8
sc1036:59984–60178	+		59984	60178	0.0
	–		60508	59764	99.9

Continued on next page

Table S3, continued

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc1036:60993–61850	+		60993	61850	27.5
	–		61960	60747	100.0
sc1036:62937–63102	+		62122	63102	100.0
	–		64214	62937	97.7
sc1038:24981–25209	+		24891	25259	67.2
	–		25209	24981	52.8
sc1038:40735–41170	+		40735	41190	66.9
	–		41177	40347	99.9
sc1038:78785–78877	+		78524	78887	100.0
	–		80593	78785	99.2
sc1038:99617–99774	+		98328	99784	99.7
	–		100673	99617	99.0
sc1038:140777–140824	+		140757	140824	36.8
	–		140824	140767	6.3
sc1040:36238–36577	+	*NCS52*	36227	36660	99.1
	–		36577	36238	0.0
sc1040:37256–37420	+		37206	37490	84.2
	–		37420	37246	3.6
sc1040:123627–123701	+		122977	123721	100.0
	–		123711	123627	98.8
sc1040:180260–180348	+		180260	180348	55.1
	–		180350	180260	44.9
sc1040:195062–195337	+		195012	195337	61.3
	–		195337	195062	0.0
sc1041:89645–89736	+		88805	89746	100.0
	–		90562	89645	94.1
sc1041:101767–102804	+		101687	102814	100.0
	–		102814	101767	20.3
sc1043:4383–4807	+		3303	4892	99.9
	–		4807	4383	23.8
sc1043:14277–14528	+		14277	14538	21.0
	–		14528	14097	65.0
sc1043:45378–45481	+		45378	45602	82.6
	–		45481	45378	0.0
sc1043:45640–45767	+		45610	45807	94.4
	–		45767	45640	0.0
sc1044:46–774	+		26	784	97.2
	–	*NCS37*	874	46	65.3
sc1044:66037–66426	+		65601	66866	100.0
	–		66436	66037	18.8
sc1044:70189–71131	+		70189	71131	85.6
	–		71131	70139	42.9
sc1044:71474–71561	+		71414	71673	92.2
	–		71561	71474	28.4
sc1044:73216–73441	+		73216	73551	91.1
	–		73441	73216	11.1
sc1045:5356–5590	+		5186	6146	100.0
	–		5590	5316	27.3
sc1045:11221–11566	+		11221	11566	0.0
	–		12176	10471	100.0
sc1045:72660–72730	+		72090	72770	100.0
	–		72730	72660	69.0
sc1045:111842–111982	+		111842	111982	0.0
	–		111982	111842	0.0
sc1046:7429–7543	+		7179	7563	100.0
	–		7543	7429	41.7
sc1047:46135–46410	+		45041	46452	100.0
	–		47271	46135	90.9
sc1048:23852–24038	+	*NCS48*	23730	24042	100.0
	–		24038	23852	0.0
sc1050:34744–35601	+		34744	35601	2.9
	–	*NCS22*	35601	34714	86.8
sc1050:87452–87670	+		87452	87761	85.5
	–		87700	87452	20.1

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc1053:6186–6792	+		6186	6792	0.0
	–		6822	6146	69.1
sc1054:27472–27504	+	*NCS29*	27372	27524	98.0
	–		27504	27472	78.8
sc1055:50052–50551	+		50036	50551	84.1
	–		51125	50052	92.6
sc1055:78975–79463	+		78975	79463	17.8
	–		79521	78899	100.0
sc1055:81055–83050	+		81055	83050	36.7
	–	*NCS20*	83056	81055	99.6
sc1055:138817–138892	+		138387	138982	100.0
	–		138932	138777	79.5
sc1055:147036–147158	+	*NCS66*	146756	147370	97.4
	–		147158	147016	35.0
sc1055:156208–156447	+		156208	156447	10.4
	–		156447	156198	12.5
sc1055:158094–158319	+		157931	158761	98.9
	–		158319	158084	10.6
sc1055:162746–162961	+		162726	162973	85.4
	–		162961	162746	23.1
sc1056:17367–17625	+		16817	17825	96.2
	–		17695	17367	43.5
sc1056:31494–31606	+		29434	31606	100.0
	–		32656	31494	99.6
sc1056:145888–146432	+		145828	146432	26.7
	–		146457	145875	100.0
sc1056:165453–165750	+	*NCS18*	165443	165750	57.5
	–	*NCS7*	167232	165323	99.8
sc1056:166532–168895	+		166532	169001	88.0
	–	*NCS6*	168905	165322	96.0
sc1056:174775–174961	+		174035	174961	90.2
	–		175389	174775	99.5
sc1057:22742–23397	+		22712	23407	64.2
	–		23397	22102	93.4
sc1057:36897–37626	+		36897	37626	6.8
	–		37727	36797	66.7
sc1058:19613–20488	+		19613	20488	28.9
	–		20488	19613	12.0
sc1058:20682–21452	+		20682	21462	27.3
	–		21525	20551	90.4
sc1058:45831–46126	+		45226	46715	100.0
	–		46126	45831	0.0
sc1058:63376–63791	+		63376	63821	31.6
	–		63821	63366	95.2
sc1058:91631–92431	+		91541	92451	99.4
	–		93011	91611	90.8
sc1058:181315–181403	+		180232	181413	99.9
	–		182233	181295	99.7
sc1058:182450–182819	+		182440	182819	13.2
	–		183209	182368	99.9
sc1058:182991–183161	+		182991	183171	27.6
	–		183208	182368	100.0
sc1058:213594–214166	+		213574	214186	57.7
	–	*NCS56*	214266	213584	93.9
sc1059:25–2594	+		25	2594	77.5
	–		2607	15	81.9
sc1059:2851–3306	+		2851	3306	0.0
	–		3307	2571	91.3
sc1059:102338–102697	+		102306	102712	100.0
	–		102697	102338	0.0
sc1059:112648–112755	+		112418	112756	99.7
	–		112765	112648	75.0
sc1059:144909–146497	+		144909	146497	15.9
	–	*NCS12*	146507	144759	99.2

	Inferred Precursor Transcript				
Cluster coordinates	Strand	Name	Start	End	Coverage (%)
sc1060:25565–25673	+		24400	25673	98.5
	–		25993	25555	100.0
sc1060:39108–39165	+		38548	39200	100.0
	–		39751	39088	98.9
sc1060:42579–42718	+		42498	42721	100.0
	–		42976	42569	96.2
sc1060:52921–53199	+		52415	53199	94.7
	–		53199	52921	38.7
sc1061:40596–40933	+		40143	40933	100.0
	–		40943	40596	38.8
sc1061:41318–41547	+		41308	41547	10.4
	–		41567	40908	100.0
sc1061:133662–133857	+		133662	133867	4.1
	–		133917	132348	100.0
sc1063:2701–2992	+		1104	2992	93.4
	–	*NCS73*	3112	2701	96.0
sc1063:3222–3443	+		3202	3523	82.6
	–		3443	3222	22.5
sc1063:14927–15332	+		14927	15332	9.9
	–		15818	14827	100.0
sc1063:20875–20976	+		18455	21016	99.8
	–		21276	20875	86.3
sc1063:56615–56692	+		56605	56702	89.8
	–		56692	56615	32.1
sc1063:57039–57172	+		57019	57282	83.7
	–		57172	57039	18.7
sc1063:58684–58771	+		58624	58783	94.9
	–		58771	58684	28.4
sc2000:4199–4861	+		4147	4911	100.0
	–	*NCS63*	5631	4199	71.8
sc2000:53765–53901	+		53765	53901	18.2
	–		53901	53765	0.0
sc2000:86459–86528	+		86459	86742	87.5
	–		86528	86459	0.0
sc2000:90080–90252	+		90080	90258	42.5
	–		90258	90080	49.7
sc2001:3713–4119	+		3673	4149	41.7
	–	*NCS50*	4279	3643	68.1
sc2002:133190–133467	+		133020	135583	100.0
	–		133467	133190	0.0
sc2002:154632–154877	+		154626	154883	100.0
	–	*NCS78*	155142	154632	98.2
sc2003:115485–115693	+		115485	115733	30.1
	–		115693	115455	65.7
sc2003:132090–133685	+		131860	133862	99.7
	–		133715	132090	26.9
sc2003:135010–135312	+		134440	135332	100.0
	–		135312	135010	8.3
sc2003:144264–144336	+		143824	144357	100.0
	–		144449	144242	100.0
sc2003:227241–227476	+	*NCS46*	227221	227716	98.4
	–		227486	227241	39.8
sc2003:260941–261041	+		260491	261071	99.3
	–		262267	260881	99.9

Table S4. mRNA-Seq analysis of wild-type (WT) and RNAi-mutant strains. Each tag was comprised of the first 25 nt of a 36-nt Illumina read.

	WT_1	WT_2	Δ*ago1* _1	Δ*ago1* _2	Δ*dcr1* _1	Δ*dcr1* _2	Total
Sequencing							
Total reads (tags)	5,237,134	5,710,767	5,469,626	5,672,984	5,481,666	5,873,485	33,445,662
Unique tags	2,362,087	2,355,724	1,792,636	2,166,169	2,079,539	2,128,205	12,884,360
Mapping of tags							
Genome-matching tags	3,913,229	4,594,533	4,256,197	4,682,941	4,609,746	4,972,396	27,029,042
% of total tags	*74.7*	*80.5*	*77.8*	*82.5*	*84.1*	*84.7*	*80.8*
Unique genome-matching tags	1,239,480	1,439,462	905,718	1,367,051	1,367,339	1,416,775	7,735,825
% of total tags	*52.5*	*61.1*	*50.5*	*63.1*	*65.8*	*66.6*	*60.0*
Total genomic hits	1,283,574	1,487,146	943,739	1,415,491	1,415,527	1,467,372	8,012,849
Analysis							
rRNA tags	490,038	562,592	699,948	615,593	613,841	755,697	3,737,709
% of genome-matching tags	*12.5*	*12.2*	*16.4*	*13.1*	*13.3*	*15.2*	*13.8*
tRNA tags	559	566	738	604	576	722	3,765
% of genome-matching tags	*<0.1*	*<0.1*	*<0.1*	*<0.1*	*<0.1*	*<0.1*	*<0.1*
Sense annotated ORF tags	2,949,357	3,426,898	3,167,181	3,518,675	3,487,311	3,705,506	20,254,928
% of genome-matching tags	*75.4*	*74.6*	*74.4*	*75.1*	*75.7*	*74.5*	*74.9*
Antisense annotated ORF tags	74,307	84,847	102,596	92,149	94,662	115,526	564,087
% of genome-matching tags	*1.9*	*1.8*	*2.4*	*2.0*	*2.1*	*2.3*	*2.1*

Table S6. Number of colonies obtained upon transformation of each strain with the plasmid indicated, labeled as in figure S11A. Three independent transformations are shown (summed in figure S11A).

Plasmid	Transformation 1			Transformation 2			Transformation 3		
	WT	Δ*ago1*	Δ*dcr1*	WT	Δ*ago1*	Δ*dcr1*	WT	Δ*ago1*	Δ*dcr1*
CEN	66	0	0	186	5	8	48	0	2
2μ	77	9	6	10	0	1	18	2	0
2μ Ago1	106	5	20	98	0	0	21	2	5
2μ Ago1, induced	68	44	3	78	26	0	23	26	1
2μ Dcr1	95	9	4	51	0	3	23	0	4
2μ Dcr1, induced	58	2	70	48	0	24	36	0	33

Table S7. Strains used and generated in this study.

Strain	Genotype	Species	Reference
FY45	*MAT***a***/MATα*	*S. cerevisiae* S288C	(*30*)
F2035	*MAT***a***/MATα*	*S. bayanus* MCYC623	NCYC
KpolWT	Wild-type	*K. polysporus* DSM70294	(*31*)
Can14	*MAT***a***/MATα*	*C. albicans* SC5314	(*32*)
F2037	Wild-type	*S. castellii* CBS4309	NCYC
Y235	*MAT***a***/MATα ura3-1/ura3-1*	*S. castellii* CBS4310	(*33*)
DPB002	*MAT***a***/MATα ura3-1/ura3-1 ago1Δ::HYG/ago1Δ::HYG*	*S. castellii* CBS4310	This study
DPB003	*MAT***a***/MATα ura3-1/ura3-1 dcr1Δ::loxP-KanMX6-loxP/dcr1Δ::loxP-KanMX6-loxP*	*S. castellii* CBS4310	This study
DPB004	*MAT***a** *ura3-1 hoΔ*	*S. castellii* CBS4310	This study
DPB006	*MAT***a** *ura3-1 hoΔ ago1Δ::HYG*	*S. castellii* CBS4310	This study
DPB008	*MAT***a** *ura3-1 hoΔ dcr1Δ::loxP-KanMX6-loxP*	*S. castellii* CBS4310	This study
DPB005	*MATα ura3-1 hoΔ*	*S. castellii* CBS4310	This study
DPB007	*MATα ura3-1 hoΔ ago1Δ::HYG*	*S. castellii* CBS4310	This study
DPB009	*MATα ura3-1 hoΔ dcr1Δ::loxP-KanMX6-loxP*	*S. castellii* CBS4310	This study
DPB220	*MATα ura3-1 hoΔ* $Flag_3$*-AGO1*	*S. castellii* CBS4310	This study
DPB313	*MATα ura3-1 hoΔ ago1Δ::HYG*	*S. castellii* CBS4310	This study
DPB318	*MATα ura3-1 hoΔ dcr1Δ*	*S. castellii* CBS4310	This study
DPB314	*MATα hoΔ ura3::EGFP(S65T)-KanMX6*	*S. castellii* CBS4310	This study
DPB317	*MATα hoΔ ago1Δ::HYG ura3::EGFP(S65T)-KanMX6*	*S. castellii* CBS4310	This study
DPB321	*MATα hoΔ dcr1Δ ura3::EGFP(S65T)-KanMX6*	*S. castellii* CBS4310	This study
DPB331	*MATα hoΔ ura3::EGFP(S65T)-KanMX6 c633::pIp*	*S. castellii* CBS4310	This study
DPB332	*MATα hoΔ ura3::EGFP(S65T)-KanMX6 c633::pIp-weakSC_GFP*	*S. castellii* CBS4310	This study
DPB333	*MATα hoΔ ura3::EGFP(S65T)-KanMX6 c633::pIp-strongSC_GFP*	*S. castellii* CBS4310	This study
DPB334	*MATα hoΔ ago1Δ::HYG ura3::EGFP(S65T)-KanMX6 c633::pIp*	*S. castellii* CBS4310	This study
DPB335	*MATα hoΔ ago1Δ::HYG ura3::EGFP(S65T)-KanMX6 c633::pIp-weakSC_GFP*	*S. castellii* CBS4310	This study
DPB336	*MATα hoΔ ago1Δ::HYG ura3::EGFP(S65T)-KanMX6 c633::pIp-strongSC_GFP*	*S. castellii* CBS4310	This study
DPB337	*MATα hoΔ dcr1Δ ura3::EGFP(S65T)-KanMX6 c633::pIp*	*S. castellii* CBS4310	This study
DPB338	*MATα hoΔ dcr1Δ ura3::EGFP(S65T)-KanMX6 c633::pIp-weakSC_GFP*	*S. castellii* CBS4310	This study
DPB339	*MATα hoΔ dcr1Δ ura3::EGFP(S65T)-KanMX6 c633::pIp-strongSC_GFP*	*S. castellii* CBS4310	This study
F2005	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	*S. cerevisiae* BY4742	(*30*)
L4718	*MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	(*34*)
DPB249	*MATα leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	This study
DPB250	*MATα leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-weakSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB251	*MATα leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB255	*MATα LEU2::pTEF-Dcr1 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	This study
DPB256	*MATα LEU2::pTEF-Dcr1 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-weakSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB257	*MATα LEU2::pTEF-Dcr1 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB258	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	This study
DPB259	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-weakSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB260	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae* W303-1B	This study
DPB271	*MATα leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	This study
DPB272	*MATα leu2-3,112 trp1-1 can1-100 ade2-1 HIS3::pGAL1-hpSC_URA3*	*S. cerevisiae* W303-1B	This study
DPB275	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ade2-1 his3-11,15*	*S. cerevisiae* W303-1B	This study
DPB276	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ade2-1 HIS3::pGAL1-hpSC_URA3*	*S. cerevisiae* W303-1B	This study

Table S8. Plasmids generated in this study.

Plasmid	Description
pYES2.1-Ago1	2-micron plasmid, *S. castellii AGO1* under *GAL1* promoter
pYES2.1-Dcr1	2-micron plasmid, *S. castellii DCR1* under *GAL1* promoter
pYES2.1-GFP	2-micron plasmid, *GFP* under *GAL1* promoter
pET101-Dcr1	*E. coli* Dcr1 expression plasmid
pIp	*S. castellii* integrating plasmid, empty
pIp-weakSC_GFP	*S. castellii* integrating plasmid, weak *GFP* silencing construct under *S. cerevisiae GAL1* promoter
pIp-strongSC_GFP	*S. castellii* integrating plasmid, strong *GFP* silencing construct under *S. cerevisiae GAL1* promoter
pRS404-P_{TEF}-Ago1	*S. cerevisiae* integrating plasmid, *S. castellii AGO1* under *TEF* promoter
pRS405-P_{TEF}-Dcr1	*S. cerevisiae* integrating plasmid, *S. castellii DCR1* under *TEF* promoter
pRS403-P_{GAL1}-weakSC_GFP	*S. cerevisiae* integrating plasmid, weak *GFP* silencing construct under *GAL1* promoter
pRS403-P_{GAL1}-strongSC_GFP	*S. cerevisiae* integrating plasmid, strong *GFP* silencing construct under *GAL1* promoter
pRS403-P_{GAL1}-hpSC_URA3	*S. cerevisiae* integrating plasmid, hairpin *URA3* silencing construct under *GAL1* promoter

Supplemental References

S1. E. Astromskas, M. Cohn, *Yeast* **24**, 499 (2007).
S2. U. Guldener, S. Heck, T. Fielder, J. Beinhauer, J. H. Hegemann, *Nucleic Acids Res* **24**, 2519 (1996).
S3. M. D. Krawchuk, W. P. Wahls, *Yeast* **15**, 1419 (1999).
S4. A. L. Goldstein, J. H. McCusker, *Yeast* **15**, 1541 (1999).
S5. M. S. Longtine *et al.*, *Yeast* **14**, 953 (1998).
S6. R. D. Gietz, R. H. Schiestl, *Nat Protoc* **2**, 31 (2007).
S7. A. Sigova, N. Rhind, P. D. Zamore, *Genes Dev* **18**, 2359 (2004).
S8. D. Mumberg, R. Muller, M. Funk, *Gene* **156**, 119 (1995).
S9. R. S. Sikorski, P. Hieter, *Genetics* **122**, 19 (1989).
S10. G. S. Pall, C. Codony-Servat, J. Byrne, L. Ritchie, A. Hamilton, *Nucleic Acids Res* **35**, e60 (2007).
S11. C. S. Hoffman, F. Winston, *Gene* **57**, 267 (1987).
S12. D. J. Garfinkel, M. F. Mastrangelo, N. J. Sanders, B. K. Shafer, J. N. Strathern, *Genetics* **120**, 95 (1988).
S13. M. J. Curcio, D. J. Garfinkel, *Proc Natl Acad Sci U S A* **88**, 936 (1991).
S14. S. E. Adams *et al.*, *Cell* **49**, 111 (1987).
S15. S. D. Youngren, J. D. Boeke, N. J. Sanders, D. J. Garfinkel, *Mol Cell Biol* **8**, 1421 (1988).
S16. A. Grimson *et al.*, *Nature* **455**, 1193 (2008).
S17. J. L. Gordon, K. P. Byrne, K. H. Wolfe, *PLoS Genet* **5**, e1000485 (2009).
S18. C. Neuveglise, H. Feldmann, E. Bon, C. Gaillardin, S. Casaregola, *Genome Res* **12**, 930 (2002).
S19. C. Llorens, R. Futami, D. Bezemer, A. Moya, *Nucleic Acids Res* **36**, D38 (2008).
S20. E. R. Havecker, X. Gao, D. F. Voytas, *Genome Biol* **5**, 225 (2004).
S21. Y. Gelfand, A. Rodriguez, G. Benson, *Nucleic Acids Res* **35**, D80 (2007).
S22. J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, *Proc Natl Acad Sci U S A* **95**, 5857 (1998).
S23. I. Letunic, T. Doerks, P. Bork, *Nucleic Acids Res* **37**, D229 (2009).
S24. C. Notredame, D. G. Higgins, J. Heringa, *J Mol Biol* **302**, 205 (2000).
S25. J. Soding, A. Biegert, A. N. Lupas, *Nucleic Acids Res* **33**, W244 (2005).
S26. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, *Science* **324**, 218 (2009).
S27. U. Nagalakshmi *et al.*, *Science* **320**, 1344 (2008).
S28. J. Houseley, K. Kotovic, A. El Hage, D. Tollervey, *Embo J* **26**, 4996 (2007).
S29. T. J. Goodwin, D. E. Dalle Nogare, M. I. Butler, R. T. Poulter, *Yeast* **20**, 493 (2003).
S30. C. B. Brachmann *et al.*, *Yeast* **14**, 115 (1998).
S31. D. R. Scannell *et al.*, *Proc Natl Acad Sci U S A* **104**, 8397 (2007).
S32. W. A. Fonzi, M. Y. Irwin, *Genetics* **134**, 717 (1993).
S33. R. F. Petersen *et al.*, *J Mol Biol* **318**, 627 (2002).
S34. B. J. Thomas, R. Rothstein, *Cell* **56**, 619 (1989).

Chapter 3

The Inside-Out Mechanism of Dicers from Budding Yeasts

David E. Weinberg[1,2,3,*], Kotaro Nakanishi[4,*], Dinshaw J. Patel[4], David P. Bartel[1,2,3]

[1] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142
[2] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139
[3] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[4] Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065

* These authors contributed equally to this work.

D.E.W. performed the biochemical experiments under the supervision of D.P.B. K.N. performed the structural experiments under the supervision of D.J.P. All authors designed the study and wrote the manuscript.

Published as:
Weinberg DE*, Nakanishi K*, Patel DJ, Bartel DP (2011). The inside-out mechanism of Dicers from budding yeasts. *Cell* 146:262–76.
** indicates equal contributions*

The Dicer ribonuclease III (RNase III) enzymes process long double-stranded RNA (dsRNA) into small interfering RNAs (siRNAs) that direct RNA interference. Here, we describe the structure and activity of a catalytically active fragment of *Kluyveromyces polysporus* Dcr1, which represents the noncanonical Dicers found in budding yeasts. The crystal structure revealed a homodimer resembling that of bacterial RNase III but extended by a unique N-terminal domain, and it identified additional catalytic residues conserved throughout eukaryotic RNase III enzymes. Biochemical analyses showed that Dcr1 dimers bind cooperatively along the dsRNA substrate such that the distance between consecutive active sites determines the length of the siRNA products. Thus, unlike canonical Dicers, which successively remove siRNA duplexes from the dsRNA termini, budding-yeast Dicers initiate processing in the interior and work outward. The distinct mechanism of budding-yeast Dicers establishes a paradigm for natural molecular rulers and imparts substrate preferences with ramifications for biological function.

RNA interference (RNAi) is a gene-silencing pathway triggered by double-stranded RNA (dsRNA) (Meister and Tuschl, 2004; Malone and Hannon, 2009). In this pathway, the RNase III enzyme Dicer first converts long dsRNA into 21–25 nucleotide (nt) small interfering RNA (siRNA) duplexes bearing 2 nt 3′ overhangs, a defining characteristic of RNase III cleavage products (MacRae and Doudna, 2007). The siRNA duplexes are then loaded into the effector protein Argonaute, and after one strand is discarded the remaining single-stranded RNA (ssRNA) pairs to RNA substrates, thereby guiding Argonaute to cleave these target transcripts (Tomari and Zamore, 2005; Wang et al., 2009b). Although RNAi or related silencing pathways are found in most eukaryotes, they are missing in *Saccharomyces cerevisiae*, which lacks both

Dicer and Argonaute homologs. Nonetheless, other budding-yeast species, including *Saccharomyces castellii* and *Kluyveromyces polysporus*, contain a noncanonical Dicer, Dcr1, which fuels the Argonaute-mediated RNAi pathway in these organisms (Drinnenberg et al., 2009).

Dicer enzymes must generate siRNAs of a length compatible with loading into Argonaute. Canonical Dicers measure from the end of the dsRNA to determine product length, with siRNA length equaling the distance spanning the terminus-binding PAZ domain and RNase III active sites (Zhang et al., 2004; Macrae et al., 2006). The noncanonical Dicers of budding yeast have a very different domain architecture that lacks a PAZ domain (Figure 1A), raising the question of how they cleave dsRNA at precise intervals. Here, we show how budding-yeast Dicers generate siRNAs of the correct length. Unlike canonical Dicers, which successively remove siRNA duplexes from the dsRNA termini, budding-yeast Dicers start in the interior and work outward. The distinct mechanisms of canonical and noncanonical Dicers confer different constraints on the types of substrates processed into guide RNAs and thus have important implications for RNAi function.

Purified Dcr1 and Dcr1ΔC retain precise cleavage activity

One way that budding-yeast Dicer might have cleaved at precise intervals without a PAZ domain could have been through association with a protein cofactor that substituted for the PAZ domain. Arguing against this possibility, recombinant *K. polysporus* Dcr1 purified from *E. coli* converted body-labeled dsRNA into 23 nt products (Figure 1B). To facilitate subsequent biochemical and structural analyses, we identified a fragment of *K. polysporus* Dcr1 that could be purified from contaminating RNA (Figure S1A available online) and had enhanced stability and solubility. Spanning Ser15 to Glu355, this fragment lacked the extreme N terminus and the C-terminal dsRBD (dsRBD2) and is referred to as Dcr1ΔC (Figure 1A). It had robust activity and generated 23 nt products indistinguishable from those of full-length enzyme or *K. polysporus* extracts (Figure 1B). Despite lacking one dsRBD, Dcr1ΔC retained specificity for dsRNA over ssRNA substrates (Figure S1B). In contrast, a fragment lacking both dsRBDs (Dcr1Δ2d) cleaved much less efficiently, failed to preferentially generate 23 nt products, and did not discriminate between dsRNA and ssRNA (Figure S1C). The requirement of dsRBD1 for dsRNA-specific cleavage activity was consistent with its requirement for dsRNA binding, as shown in a gel-shift assay (Figure S1D).

Sequencing of 18–30 nt RNAs isolated from processing reactions performed with full-length Dcr1 and Dcr1ΔC (Table S1) showed that products of both enzymes were predominantly 23 nt and displayed little 5′ nucleotide bias (Figure 1C). The abundances of 23 nt species in the two libraries were strongly correlated ($R^2 = 0.84$), which showed that Dcr1ΔC faithfully recapitulated the siRNA-generating activity of full-length Dcr1 (Figure 1D). The length homogeneity of in vitro products exceeded that of endogenous siRNAs in *K. polysporus* (Drinnenberg et al., 2009) (Figure 1C). Perhaps length heterogeneity observed in vivo reflects subsequent events, such as 3′ end trimming of siRNAs that have been loaded into Argonaute (Halic and Moazed, 2010). In addition, the absence of a 5′ uridine bias among siRNAs generated in vitro suggested that the bias observed in vivo might arise from the binding specificity of Argonaute (Wang et al., 2009b; Frank et al., 2010).

Although dsRBD2 was dispensable for size-specific siRNA generation in vitro (Figures 1B–1D), its strict conservation among Dcr1 proteins and absence from Rnt1 proteins (RNase III enzymes involved in yeast ribosome biogenesis, Figure 1A) prompted us to examine its importance for siRNA accumulation in vivo using *S. castellii*, a species that is closely related to *K. polysporus* but has more tools available for molecular-genetic manipulation (Drinnenberg et al., 2009). The endogenous copy of *S. castellii* Dcr1 was replaced with *S. castellii* Dcr1ΔC expressed from the native promoter. Although extracts made from yeast expressing full-length Dcr1 and Dcr1ΔC had comparable in vitro dicing activity, endogenous siRNAs failed to accumulate in the Dcr1ΔC strain (Figure 1E). The absence of siRNAs in this strain could not be attributed to either lack of the extreme N terminus or lower expression, as siRNA accumulation in vivo was not rescued by either restoration of the native N terminus (data not shown) or overexpression of Dcr1ΔC in a Δ*dcr1* strain (Figure 1E). In addition to size-specific cleavage, Dcr1 function in vivo might also involve appropriate protein localization, protection of product siRNAs from nucleases, and interaction with other proteins such as Argonaute. The C-terminal dsRBD is presumably involved in one or more of these additional aspects of Dcr1 function.

Crystal structure of Dcr1ΔC

Having found that Dcr1ΔC recapitulates the siRNA-generating activity of the full-length enzyme, we determined its crystal structure at 2.3 Å resolution (Figure 2A and Table S2). Because RNase III domains work in pairs (MacRae and Doudna, 2007), Dcr1 has been proposed to act as a homodimer (Drinnenberg et al., 2009). Gel filtration and glutaraldehyde crosslinking confirmed that Dcr1ΔC behaves as a homodimer in solution (Figures S2A and S2C). The crystal structure had two homodimers in the asymmetric unit (Figure S2D), such that all four N-terminal domains (NTDs) and RNase III domains had well-defined electron density, whereas only one of the four dsRBD1 domains had traceable density, presumably because it had stabilizing crystal-packing interactions (Figure 2B, left panel). Dcr1ΔC forms a compact homodimer through the

NTD and RNase III domains (Figure 2A), with the single observed dsRBD1, which exhibited variable density (Figure 2B), connected to the RNase III domain by a flexible linker. The buried surface within the dimer is 1,850 Å^2, including 842 Å^2 between the NTDs.

The NTD has a novel fold composed of five α helices and forms a symmetric homodimer that packs against the RNase III domain homodimer (Figure 2A). Although a DALI search for structural similarity (Holm and Sander, 1993) failed to identify structures with significant homology to the NTD, sequence comparison of Dcr1 and Rnt1 N-terminal regions revealed a portion of Rnt1 with homology to the Dcr1 NTD (Figure 1A and Figure S2E), suggesting that both adopt a similar structure and have a common function and evolutionary origin. Indeed, the homodimeric nature of the Dcr1 NTD is consistent with the finding that the *S. cerevisiae* Rnt1 N-terminal region contains a dimerization signal (Lamontagne et al., 2000). Despite contributing an additional dimerization interface to the Dcr1 homodimer (Figure 2A), the NTD was not absolutely required for either dimerization (Figure S2F) or siRNA generation (Figures S1E and S1F), observations analogous to those for Rnt1 (Lamontagne et al., 2000).

The structure of the homodimeric Dcr1ΔC RNase III domain (Figure 2A) resembles structures of the homodimeric bacterial RNase III domain and the intramolecular heterodimeric RNase III domains of the canonical Dicer from *Giardia intestinalis* (*Gi*Dicer) (Gan et al., 2006; Macrae et al., 2006). One difference is that the Dcr1ΔC RNase III domains swap their terminal α helices with each other (Figure 2A), which would presumably further strengthen the very stable interaction observed for RNase III homodimers (Meng and Nicholson, 2008). Our crystal structure of an active-site mutant (E224Q) of Dcr1Δ2d, which diffracted to higher (1.97 Å) resolution (Figure 2C, Table S2, and Figure S2B), confirmed this domain swap, as illustrated in a comparison with the *Aquifex aeolicus* RNase III (*Aa*RNase III) homodimer

(Figures 2D and 2E). Another difference from the structures of *Aa*RNase III and *Gi*Dicer is that Dcr1ΔC has two long loops that correspond to the most variable regions in RNase III alignments (Figures 2A and 3A). We refer to the loops between α6 and α7 and between α11 and α12 as variable loops 1 and 2 (VL-1 and VL-2), respectively. α7, to which VL-1 is connected, is an additional helix that is not observed in other RNase III enzyme structures.

The disorder of the dsRBD1s in the Dcr1ΔC structure was consistent with reports of flexibility in the orientations between RNase III domains and dsRBDs in the absence of dsRNA (Akey and Berger, 2005). Nonetheless, the robust electron density observed for one dsRBD1 permitted confident determination of its structure (Figure 2B). Dcr1 dsRBD1 has the canonical α-β-β-β-α topology, with the pair of α helices packed against an antiparallel β sheet (Figure 2B). Among previously solved dsRBD structures, Dcr1 dsRBD1 most closely resembles Rnt1 dsRBD, in that these two dsRBDs are the only ones that include an α-helical extension (α17) beyond the canonical dsRBD fold (Leulliot et al., 2004). However, Dcr1 has not conserved the identities of residues that confer to Rnt1 a preference for AGNN tetraloops (Wu et al., 2004), which explains an absence of tetraloop specificity for Dcr1 (Figure S2G). In addition, the β1-β2 loop is especially short in Dcr1 dsRBD1 and thus would not be able to insert into the minor groove, which would presumably weaken affinity between this dsRBD and dsRNA (Figure S2H).

Sequence alignment of RNase III domains had suggested a close evolutionary relationship between Rnt1 and Dcr1 (Drinnenberg et al., 2009). Our results extend this relationship to two additional domains, showing that Rnt1 and Dcr1 have similar NTDs that expand the homodimer interface and that Dcr1 dsRBD1 shares many features with Rnt1 dsRBD but has lost otherwise conserved elements required for sequence specificity.

Metal-ion coordination in the active site

The RNase III domains of *Aa*RNase III and *Gi*Dicer form intermolecular and intramolecular dimers, respectively, to position a pair of active sites ∼20 Å apart, which accommodates the width of the dsRNA minor groove with a 2 nt offset (Gan et al., 2006; Macrae et al., 2006). Within each active site, the most prominent metal-binding site (M1) is surrounded by four acidic residues (two Glu and two Asp), which are strictly conserved among RNase III enzymes and required for efficient catalysis (Takeshita et al., 2007). The RNase III domains of budding-yeast Dicers contain the four conserved acidic residues: Glu147, Asp151, Asp221, and Glu224 (Figure 3A). We observed clear octahedron electron density in the middle of these residues of Dcr1ΔC, indicating that each active site contained a metal ion (Figure 3B). Because the crystallization buffer contained 20 mM Mg^{2+}, the ion presumably is Mg^{2+}. The 20.8 Å distance between the two metals within each Dcr1ΔC dimer (Figure 3B) is consistent with that of other RNase III enzymes (Takeshita et al., 2007).

In the structure of *Aa*RNase III (Blaszczyk et al., 2004), the metal ion at the M1 site is six-coordinated to Glu40, Asp107, and Glu110 (corresponding to Glu147, Asp221, and Glu224 of *K. polysporus* Dcr1, respectively) and three water molecules (w1–w3) (Figure 3C). The M1 site of Dcr1 is six-coordinated in a similar manner, with Glu147 and Glu224 participating in metal-ion binding in the same way as the corresponding residues in *Aa*RNase III (Figures 3B and 3C). In both structures, the metal-coordinated w3 is also hydrogen bonded to Glu110/224 (Figure 3C). Previous biochemical analyses of bacterial RNase III mutants demonstrated that Glu110 binds two Mg^{2+} ions and is essential for dsRNA cleavage, whereas the other active-site residues are not absolutely required for cleavage (Sun and Nicholson, 2001; Sun et al., 2004). As observed for the bacterial enzymes, replacement of Glu224 with Gln completely eliminated Dcr1ΔC activity, whereas a substitution of both Glu147 and Asp151 (with Gln and Asn, respectively) greatly diminished (Figure 3D) but did not eliminate (Figure S3) activity. As in most other apo-RNase III structures, no electron density was observed for a second metal ion in the active sites of our structure. By analogy to *Aa*RNase III (Gan et al., 2008), Glu224 of Dcr1 might coordinate two metal ions only after substrate RNA enters the active site.

An important difference between the M1 site of Dcr1 and that of bacterial RNase III is that in Dcr1 a fourth water molecule (w4), rather than Asp221, directly coordinates the Mg^{2+} ion (Figures 3B and 3C). In both active sites, this water is positioned by Asp221 as well as Asn184. An analogous Asn residue was proposed to hydrogen bond with a water molecule in a canonical Dicer RNase IIIb domain, which lacked a metal ion (Du et al., 2008). Our structure provides a function for this interaction, indicating that the water and thus Asn184 help position an active-site metal ion. In Dcr1, a sixth residue, Lys217, also contributes to metal-ion binding. In one of the active sites Lys217 forms hydrogen bonds with both w1 and w4, whereas in the other active site Lys217 forms hydrogen bonds with w1 and Asp221, which is associated with distortion of the octahedron (Figure 3B). Structural analysis of an *Aa*RNase III-product dsRNA complex indicates that two oxygen atoms of the scissile phosphate group replace w1 and w2 in the octahedral coordination (Gan et al., 2008). By analogy, Lys217 of Dcr1 would directly recognize the scissile phosphate group during the cleavage reaction.

The residues corresponding to Asn184 and Lys217 are not conserved in bacterial RNase III enzymes but are conserved throughout eukaryotic RNase III enzymes (with the exception of Lys217 in the RNaseIIIa domain of Drosha, Figure 3A), suggesting that they have an important catalytic role in eukaryotes. To test their importance for cleavage activity, we examined effects of Ala substitutions at Asn184 or Lys217. Replacement of Asn184 reduced in vitro cleavage activity to extremely low levels (Figure 3D and Figure S3). Replacement of Lys217 also reduced activity (Figure 3D and Figure S3), as previously observed for the corresponding residue in a canonical Dicer RNaseIIIb domain (Du et al., 2008). Thus, their positions in the structure, evolutionary conservation, and sensitivity to mutation all indicate that these two residues play important roles in the active sites of eukaryotic RNase III enzymes.

Closely spaced Dcr1 dimers generate 23 nt products from internal dsRNA regions

Because the NTD abuts the RNase III domain on the side opposite that of the catalytic surface (Figure 2A), a Dcr1ΔC dimer could interact with dsRNA in a manner analogous to that observed for *Aa*RNase III (Figure 4A). When apo-Dcr1ΔC and an *Aa*RNase III-dsRNA complex (Gan et al., 2006) are structurally aligned based on their RNase III domains, the modeled dsRNA only clashes with VL-1 of Dcr1, suggesting an induced fit wherein VL-1 extends along an RNA groove (Figure 4B). The alignment also suggested an induced fit of VL-2, because the corresponding loop in *Aa*RNase III (which constitutes RNA-binding motif 4) moves upon dsRNA binding (Gan et al., 2006). By using the loops VL-1 and VL-2 protruding from the sides, Dcr1ΔC would have an interface with dsRNA larger than that of *Aa*RNase III. To evaluate the importance of VL-1 and VL-2 for Dcr1 activity, we replaced these regions in Dcr1ΔC with the structurally analogous regions from the *Gi*Dicer RNase IIIb domain (Figure S4A). The VL-2 substitution abrogated cleavage activity, and the VL-1 substitution reduced activity such that even after extended incubation only heterogeneous products were observed (Figure 4C). Gel-shift assays indicated that the mutants had greatly reduced affinities for dsRNA substrate (Figure S1D).

The structural alignment with *Aa*RNase III-dsRNA revealed that neither the NTD nor the dsRBD of Dcr1ΔC are positioned to measure the length from the end of a dsRNA molecule to the active sites (Figure 4A). Therefore, we considered the possibility that unlike canonical Dicers, which sequentially remove siRNAs from termini of dsRNA, budding-yeast Dicers might generate 23 nt products from internal regions of dsRNA. To test this possibility, we compared the products of *K. polysporus* and canonical Dicers when acting on body- and end-labeled dsRNA substrates. Dcr1ΔC processing of body-labeled dsRNA yielded a predominant 23 nt radiolabeled product, whereas processing of end-labeled dsRNA yielded mostly shorter, variable-length radiolabeled fragments, without enrichment of 23 nt product (Figure 4D). In contrast, human Dicer generated radiolabeled siRNA products from both body- and end-labeled dsRNA substrates. These results indicate that internal regions of dsRNA but not the termini are incorporated into Dcr1-generated siRNAs.

The short lengths of end-labeled products generated by Dcr1ΔC suggested that they were terminal cutoff products. The length heterogeneity of Dcr1ΔC cutoff products (Figure 4D) suggested that the initiation of dsRNA processing from within the dsRNA molecule occurs at multiple potential sites (i.e., without a dominant register). Consistent with this interpretation, products deriving from a perfectly paired duplex were only weakly phased (Figure S4B). The stronger phasing observed for endogenous siRNAs (Drinnenberg et al., 2009) can be explained by loops and bulges present in many natural Dcr1 substrates, which could bias processing toward specific registers.

To understand how Dcr1ΔC might generate 23 nt products from internal regions of dsRNA, we manually docked two Dcr1ΔC dimers at sites 23 nt apart on a dsRNA substrate. The dimers aligned near to each other but without steric hindrance in a configuration that would generate 23 nt siRNA products with 2 nt 3′ overhangs (Figure 4E). This docking model for Dcr1ΔC is remarkably similar to the crystal packing observed in the structure of *Aa*RNase III bound to a 22 nt dsRNA, in which RNase III dimers are adjacently positioned along a pseudocontinuous dsRNA generated by end-to-end stacking of short RNA helices (Figure S4C). Based on these structural observations, we hypothesized that Dcr1 dimers bind closely together along dsRNA and that the distance between adjacent active sites determines product length.

To test the first part of this hypothesis, we used glutaraldehyde crosslinking to determine whether the binding to dsRNA brings dimers sufficiently near to each other to be crosslinked into higher-order oligomers. For these experiments, we utilized a Dcr1ΔC mutant that contained an RNase III active-site substitution (E224Q) that prevents cleavage (Figure 3D) but does not affect substrate binding of *E. coli* RNase III (Sun and Nicholson, 2001). Although base-paired RNA does not react with glutaraldehyde (Hopwood, 1975), control experiments with dsDNA instead of dsRNA were used to identify any effects of double-stranded nucleic acid on crosslinking that were independent of Dicer binding. Inclusion of dsRNA shifted the distribution of species to molecular weights higher than those seen with no nucleic acid or dsDNA (Figure 4F). In lower glutaraldehyde (0.01% and 0.03%) the presence of dsRNA reduced intradimer crosslinking, as indicated by uncrosslinked monomer, but increased interdimer crosslinking, as indicated by higher molecular-weight bands. In 0.1% glutaraldehyde the addition of dsRNA caused nearly all of the protein to be crosslinked into higher-order oligomers. Larger species were observed when a 500 bp dsRNA was used compared to when a 70 bp dsRNA was used, consistent with the ability of the longer dsRNA to accommodate more dimers. Experiments following the fate of the dsRNA confirmed that the dsRNA-dependent crosslinked species represented protein-protein crosslinks rather than protein-RNA or RNA-RNA crosslinks (Figures

S4D and S4E). Together, these results showed that Dcr1ΔC dimers can bind closely together along dsRNA.

Cleavage by active sites in adjacent Dcr1 dimers generates 23 nt products

To test the second part of our hypothesis, that the distance between adjacent active sites determines product length, we changed the distance between active sites to see if this caused a corresponding change in product length. To increase the distance between functional active sites, Dcr1ΔC with an active-site mutation (E224Q) was added to reactions containing wild-type Dcr1ΔC. If multiple Dcr1 dimers bind side-by-side along dsRNA, then binding of one or more E224Q dimers between two active dimers would increase the distances between active sites by multiples of 23 nt, with corresponding increases in product length (Figure 5A, top). Indeed, increasing E224Q Dcr1ΔC concentration in the presence of constant wild-type Dcr1ΔC produced progressively longer ladders with fragment lengths increasing in the predicted 23 nt increments (Figure 5A, bottom). A 46 nt fragment was also observed in the first few seconds of processing by wild-type Dcr1 (Figure 1B), which is explained by our model as a short-lived intermediate that is a consequence of cleavage occurring in the context of multiple Dcr1 active sites spaced along the dsRNA at 23 nt intervals. Likewise, larger intermediates with lengths in multiples of 23 nt were observed in reactions containing Dcr1ΔC active-site mutants (Figure 3D and Figure S3), as expected for rare cleavage events by closely packed dimers aligned on the dsRNA substrate.

Discrete processing intermediates separated by a constant interval equivalent to the dominant siRNA product length have been observed with canonical Dicers (Zhang et al., 2002; Macrae et al., 2006). These intermediates result from sequential removal of products from the ends, and thus their lengths are dependent on substrate length (Figure 5B, top right), whereas in our model for Dcr1 the lengths of intermediates should be independent of substrate length (Figure 5B, top left). Using 65, 70, and 75 bp substrates, a mixture of wild-type and E224Q Dcr1ΔC generated a 46 nt product irrespective of substrate length (Figure 5B, bottom). In contrast, human Dicer generated intermediates with lengths that varied predictably with substrate length, consistent with sequential removal of siRNA products from the ends. Thus, budding-yeast Dicer intermediates are fundamentally different from canonical Dicer intermediates, and the ladders in Figure 5A must result from discrete increases in active-site spacing.

According to our model for siRNA generation by budding-yeast Dicer, a substrate must be able to accommodate at least two dimers to preferentially generate a 23 nt product. Consequently, the shortest dsRNA substrate that could yield siRNA products should have a length corresponding to the length of dsRNA required for efficient binding of a pair of adjacent dimers. Structural modeling of Dcr1ΔC with dsRNA indicates that a pair of dimers spans 35–40 bp, excluding the outer VL-1 and VL-2 (Figure 5C). Indeed, reactions with substrates of at least 40 bp generated a predominant 23 nt product, whereas those with shorter substrates yielded little or no enrichment for 23 nt products and instead produced a collection of ≤23 nt products, which presumably derived from binding and cleavage by a single dimer (Figure 5C). Even 23 nt siRNA-like duplexes could be processed by Dcr1 into shorter fragments, albeit inefficiently (Figure S1G). The absence of >23 nt products in reactions with 30 bp substrates indicated that cleavage by a single dimer was more efficient when at least 7 bp flanked the cleavage site. The finding that the minimum substrate length for efficient siRNA production corresponded to the footprint of a pair of dimers further supported our model for the generation of 23 nt products.

Our model for budding-yeast Dicer predicts that even a dsRNA substrate that lacks termini can be processed into 23 nt siRNA products. In contrast, canonical Dicers cannot efficiently process substrates that lack free helical ends because of their reliance on the terminus-binding PAZ domain for measurement (Zhang et al., 2002; MacRae et al., 2007). To test these predictions, we compared the abilities of *K. polysporus* Dcr1ΔC and *Drosophila melanogaster* Dcr-2 (*Dm*Dcr-2)—a canonical Dicer specifically involved in siRNA biogenesis (Cenik et al., 2011)—to process dsRNA substrates with either two, one, or zero free termini (Figure 5D, top; substrates A, B, and C, respectively). Dcr1ΔC processed all substrates into siRNA products with similar efficiency, whereas *Dm*Dcr-2 efficiently processed only the substrates with at least one free end (Figure 5D), with the residual siRNA generation observed for the substrate without a free end at least partially attributed to contamination by a substrate with free ends (Figure S5B). *Dm*Dcr-2 was also more sensitive to the geometry of the substrate termini (Figure S5C).

The terminus-independent processing by budding-yeast Dicer has implications for the generation of siRNAs in vivo because endogenous substrates of Dcr1 lack the dsRNA termini preferred by canonical Dicer and instead have long segments of single-stranded RNA extending from the duplex (Drinnenberg et al., 2009). To evaluate whether budding-yeast Dicer is uniquely suited to process these classes of substrates, we compared the abilities of Dcr1ΔC and *Dm*Dcr-2 to process three substrates: a perfect duplex; a palindromic RNA cloned from *S. castellii*, which forms a hairpin with 5′ and 3′ terminal extensions; and a sense-antisense pair comprised of a capped and polyadenylated RNA paired within the body of a longer transcript (Figure 5E; substrates D, E, and F, respectively). Dcr1ΔC efficiently processed all three substrates, whereas *Dm*Dcr-2 efficiently processed only the simple duplex (Figure 5E, bottom), suggesting that the termini of many endogenous Dcr1 substrates are not recognized by the PAZ domain. Although cofactors or preprocessing by nucleases might allow such non-ideal substrates to become processed by canonical Dicers, thereby explaining the phased siRNAs generated from hairpin precursors by *Dm*Dcr-2 in vivo (Czech et al., 2008 and Okamura et al., 2008), these intrinsic differences in substrate specificity highlight the distinct mechanisms of budding-yeast and canonical Dicers.

Slow product release and cooperative binding contribute to siRNA generation

Our model relies on a second dimer binding at the proper position before the first dimer has cleaved and released its products. This could be accomplished through cooperative binding along the length of the substrate and might be further favored by slow product release. To determine whether product release was slow relative to substrate binding and/or catalysis, we analyzed multiple-turnover kinetics. Two phases were observed, which corresponded to a pre-steady-state burst followed by slower steady-state turnover, with the amplitude of the burst increasing proportionately with the enzyme concentration (Figures 6A and 6B and Figure S6A). The simplest explanation for this behavior is that product release became rate limiting after the first round of cleavage.

To visualize the protein-bound state of siRNA products, we analyzed single-turnover reactions on a nondenaturing gel. In reactions with wild-type Dcr1ΔC that had proceeded to completion, as shown on a denaturing gel (Figure 6C, left), RNA-containing species migrated as a smear on a nondenaturing gel (Figure 6C, right). After thermal denaturation, the smear resolved into unbound 23 nt siRNA and ssRNA, suggesting that it represented protein-bound RNA. Thus, a large fraction of siRNA products remained associated with Dcr1ΔC after cleavage occurred.

Together, our results indicate that slow product release contributes to siRNA generation by enabling additional dimers to bind adjacent to previously bound dimers before the previously bound dimers release their products. Slow product release might also protect the siRNAs from further nucleolytic cleavage, and it might enable Dcr1 to escort siRNAs to Argonaute and facilitate loading, as has been reported for canonical Dicers (Lee et al., 2004 and Wang et al., 2009a).

The ability of substoichiometric amounts of Dicer to preferentially generate 23 nt products in the presence of excess dsRNA-binding sites (Figure 6A) implied that binding of Dcr1ΔC to substrates was cooperative; otherwise, Dcr1ΔC dimers would bind randomly throughout the bodies of dsRNA molecules, only rarely positioning themselves precisely 23 nt apart to generate the proper product (Figure 6D). To examine binding of Dcr1 to dsRNA substrates, we performed gel-shift assays with E224Q Dcr1ΔC. Using a 70 bp dsRNA substrate, which can accommodate up to three dimers, increasing amounts of Dicer led to appearance of a single predominant gel-shifted band (Figure 6E). Although mobility in nondenaturing gels can be difficult to interpret, the observed pattern was consistent with the all-or-none behavior expected for strong cooperativity.

Binding curves for a series of dsRNA substrates showed that longer substrates had higher affinities (Figure 6F and Figures S6B and S6C). The affinities for 70, 140, and 280 bp substrates were similar but still correlated with substrate length, an observation consistent with Dicer binding to internal regions of dsRNA, as longer substrates contain more potential binding sites. Notably, the apparent K_D for the 30 bp dsRNA—which could fully accommodate only a single dimer and did not preferentially generate 23 nt siRNA products—was ~50-fold weaker than that of the 70 bp dsRNA. The 30 and 25 bp substrates exhibited similar Dicer-binding behavior, but a

20 bp substrate failed to form significant amounts of stable complex detectable by gel shift, even when Dicer was present at concentrations as high as 1 μM (Figure 6F and Figure S6). The striking sigmoidicity of the binding isotherms for all substrates ≥25 bp provided further evidence for cooperative binding, with Hill coefficients for substrates ≥70 bp reaching a plateau of ~3 (Figure 6F). Our experiments thus identified three types of dsRNA substrates with respect to binding affinity: long substrates (≥70 bp), which could fully accommodate at least three dimers and displayed pM affinities and Hill coefficients of ~3; shorter substrates (25–30 bp), which formed sufficient interactions with a pair of dimers to exhibit nM affinities and Hill coefficients of ~2; and a 20 bp substrate, which presumably bound only a single dimer with μM affinity and a Hill coefficient of ~1. Increasing monovalent ion concentrations from 41 mM to 150 mM reduced the binding affinity for the 140 bp dsRNA by ~5-fold and reduced the Hill coefficient to 1.6 (Figure S6D). However, Dcr1ΔC still behaved cooperatively under this more physiological salt concentration, as evidenced by both a Hill coefficient that exceeded unity and the preferential generation of 23 nt products under multiple-turnover conditions (Figure S6E). We conclude that binding cooperativity contributes to siRNA generation by precisely positioning dimers 23 nt apart even at limiting Dicer concentrations (Figure 6A), and this, supplemented by slow product release, helps prevent isolated dimers from enacting non-productive cleavage.

Discussion

Our results show that budding-yeast Dicers produce siRNAs through a mechanism different from that of canonical Dicers. Instead of successively removing siRNA duplexes from the dsRNA termini, Dcr1 starts in the interior and works outward (Figure 7). This inside-out mechanism initiates with a dimer binding at an arbitrary position within the dsRNA, followed by the recruitment of additional dimers to adjacent sites. As binding propagates in both directions along the dsRNA, slow product release prevents cleavage events from disrupting maintenance of the phase. Cleavage by a collection of aligned dimers precisely generates 23 nt siRNA products paired to each other with 2 nt 3′ overhangs.

The mechanism for Dcr1-catalyzed siRNA production represents a natural example of a molecular ruler that is defined by the spacing of adjacent active sites. The concept of such a molecular ruler has been proposed but then rejected for other enzymes. The multimeric proteasome was hypothesized to generate short peptide products with a length determined by the distance between active sites (Wenzel et al., 1994), but subsequent experiments ruled out this model (Nussbaum et al., 1998). Ironically, models for product length determination based on active-site spacing were also proposed for both bacterial RNase III and canonical Dicers (Blaszczyk et al., 2001; Carmell and Hannon, 2004), but further study of these enzymes proved these models to be incorrect (Zhang et al., 2004). Nonetheless, under certain conditions, *E. coli* RNase III can process long dsRNA into ~23 nt products in vitro by using a mechanism that might resemble the inside-out mechanism described here (Xiao et al., 2009). Products of this in vitro reaction act as potent siRNAs for mammalian gene knockdown (Yang et al., 2002; Xiao et al., 2009), as do siRNAs generated by budding-yeast Dicer (Figure S7).

In canonical siRNA-generating Dicers, the helicase domain uses ATP to facilitate complete processing of a duplex into siRNAs before beginning on the next duplex (Cenik et al., 2011; Welker et al., 2011). In budding-yeast Dicers, cooperativity could facilitate complete processing without requiring such a domain. dsRBD1 and VL-1/2 are candidates for forming cooperative interactions between adjacent Dcr1 dimers bound to dsRNA. Given their roles in dsRNA binding (Figure S1D), dissecting their potential contributions to cooperativity awaits a high-resolution view of the Dcr1-dsRNA complex, which would reveal dimer-dimer interactions that might be abolished without perturbing dimer-dsRNA interactions. In addition to mediating cooperativity, dimer-dimer interactions might allosterically activate adjacent dimers for cleavage, which would further favor productive cleavage.

In the current RNase III enzyme classification, which is based on domain architecture, class I includes both bacterial RNase III and yeast Rnt1 (MacRae and Doudna, 2007). We found that despite having similar domain architectures, bacterial RNase III and yeast Rnt1/Dcr1 use distinct active-site arrangements comprising four and six residues, respectively (Figures 3A and 3C). Adding this feature to the existing classification criteria would divide RNase III enzymes into four classes more parsimonious with their evolutionary relationships: bacterial RNase III, class I; Drosha, class II; canonical Dicer, class IIIa; and yeast RNase III, including both Rnt1 and Dcr1, class IIIb (Figure 3A). Despite its closer evolutionary relationship to Dcr1, Rnt1 behaves as a molecular ruler in a manner more analogous to canonical Dicer. Just as the canonical Dicer PAZ domain binds to the 2 nt 3′ overhang of its substrate to position the RNase III active sites at a defined distance, the Rnt1 dsRBD recognizes the AGNN tetraloop of its substrate to position its active sites for precise cleavage (MacRae and Doudna, 2007). Thus, the terminus-independent measuring mechanism of Dcr1 departs from the principles operating in other class III RNase III enzymes.

The distinct mechanisms employed by canonical and budding-yeast Dicers to generate similarly sized siRNAs provide a striking example of convergent functional evolution. Although both mechanisms produce siRNAs, the canonical mechanism is more suitable for producing small RNAs that must be processed in a defined register, such as microRNAs and *trans*-acting siRNAs (Vaucheret, 2005). In contrast, the inside-out mechanism is more suitable for substrates that lack free helical ends, such as covalently closed molecules (e.g., viroids), dsRNA intermediates of rolling-circle replication, dsRNA with protected termini (e.g., viral ribonucleoproteins), and dsRNA with long single-stranded extensions, including endogenous Dcr1 substrates (Figure 5E and Figure S5D). Thus, in a budding-yeast ancestor, the presence of dsRNA species resistant to processing by the canonical Dicer might have favored the evolution of an additional RNase III enzyme able to preprocess these substrates by cutting in their interior, thereby producing suitable substrates for canonical Dicer. After this enzyme acquired features that enabled it to produce siRNAs on its own, the absence of a phased small-RNA pathway might have allowed loss of the canonical Dicer without deleterious effects, thereby explaining the replacement of the canonical Dicer in the budding-yeast clade.

Acknowledgments

We thank I.A. Drinnenberg, G. Fink, R. Fukunaga, C. Jan, R. Sauer, and D. Shechner for helpful discussions, D. Koppstein for performing mammalian gene knockdown experiments, the Whitehead Genome Technology Core for high-throughput sequencing, P. Zamore for providing *Dm*Dcr-2 enzyme, F. Roth for providing pAG416Gal-Dicer plasmid, Q. Yin in the H. Wu laboratory for assistance with light scattering experiments, the X29 beamline at Brookhaven National Laboratory, and the NE-CAT beamline at the Advanced Photon Source, Argonne National Laboratory. This work was supported by National Institutes of Health grants GM061835 (D.P.B.) and AI121493 (D.J.P.), a National Science Foundation graduate research fellowship (D.E.W.), a Human Frontier Science Program Long-term Fellowship (K.N.), and a fellowship from the Japan Society for the Promotion of Science for Research Abroad (K.N.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

Accession numbers

Small-RNA sequencing data were deposited in the Gene Expression Omnibus (GSE29168). X-ray coordinates of Dcr1ΔC and E224Q Dcr1Δ2d were deposited in the Protein Data Bank (3RV0 and 3RV1, respectively).

Methods

Protein expression and purification

Detailed methods for expression and purification of His-Sumo-tagged proteins from bacterial plasmids (Table S4) are described in the Extended Experimental Procedures.

Structure determination and refinement

Native and SeMet-substituted crystals of Dcr1ΔC were obtained by sitting-drop vapor diffusion at 20°C, and data sets were collected at Brookhaven NSLS beamline X29. Phasing was achieved via single-wavelength anomalous dispersion (SAD) with selenium anomalous signals. Native crystals of E224Q Dcr1Δ2d were grown under related conditions, data were collected at Argonne NE-CAT beamline 24-IDE, and the structure was solved by molecular replacement. For details on crystallization, structure calculation, and modeling, see the Extended Experimental Procedures.

Dicer activity assays

Processing reactions using yeast whole-cell extracts were essentially as described (Drinnenberg et al., 2009). Reactions using recombinant Dcr1 variants were in 30 mM Tris-HCl (pH 7.5), 30 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 0.1 mM EDTA. Unless indicated otherwise, these reactions were for 90 s with 30 nM purified protein and ~45 pg/µl RNA, which corresponded to ~1 nM for 70 bp dsRNA (standard single-turnover conditions). Reactions using recombinant *H. sapiens* Dicer (Genlantis) and *D. melanogaster* Dcr-2 were in 30 mM Tris-HCl (pH 6.8), 25 mM NaCl, 3 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA, and 1% glycerol, supplementing the Dcr-2 reactions with 1 mM ATP. Reactions were quenched by addition to ≥1 volume of formamide loading buffer (90% formamide, 18 mM EDTA, 0.025% sodium dodecyl sulfate, 0.1% xylene cyanol, and 0.1% bromophenol blue). RNA products were resolved by denaturing PAGE and visualized by phosphorimaging.

Dicer binding assays

Binding reactions were in 30 mM Tris-HCl (pH 7.5), 30 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA, and 5% glycerol. Reactions were incubated at room temperature (~23) for 10 min and then on ice for at least 20 min. Reactions were analyzed on native polyacrylamide gels run at 4°C, and RNA was visualized by phosphorimaging.

Glutaraldehyde crosslinking

Crosslinking was in 30 mM HEPES-NaOH (pH 7.6), 30 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA, 300 nM protein, and the indicated concentrations of nucleic acid and glutaraldehyde. After 10 min at room temperature, reactions were quenched by addition of an equal volume of 2× Laemmli sample buffer supplemented with 5% β-ME and 100 mM Tris-HCl (pH 8.0). Products were resolved by SDS-PAGE and visualized by silver staining and phosphorimaging.

Yeast manipulations

S. castellii and *K. polysporus* culture, transformations, RNA isolation, and RNA blots were essentially as described (Drinnenberg et al., 2009) with the strains and plasmids listed (Table S3 and Table S4, respectively).

Small-RNA sequencing

Single-turnover reactions were performed separately with dsRNA corresponding to fragments of mRNAs for *Renilla* luciferase and green fluorescent protein (GFP). After quenching, reactions were pooled, and total RNA was isolated by phenol extraction. Small RNAs were sequenced as described (Drinnenberg et al., 2009). For a detailed description of data analysis, see the Extended Experimental Procedures.

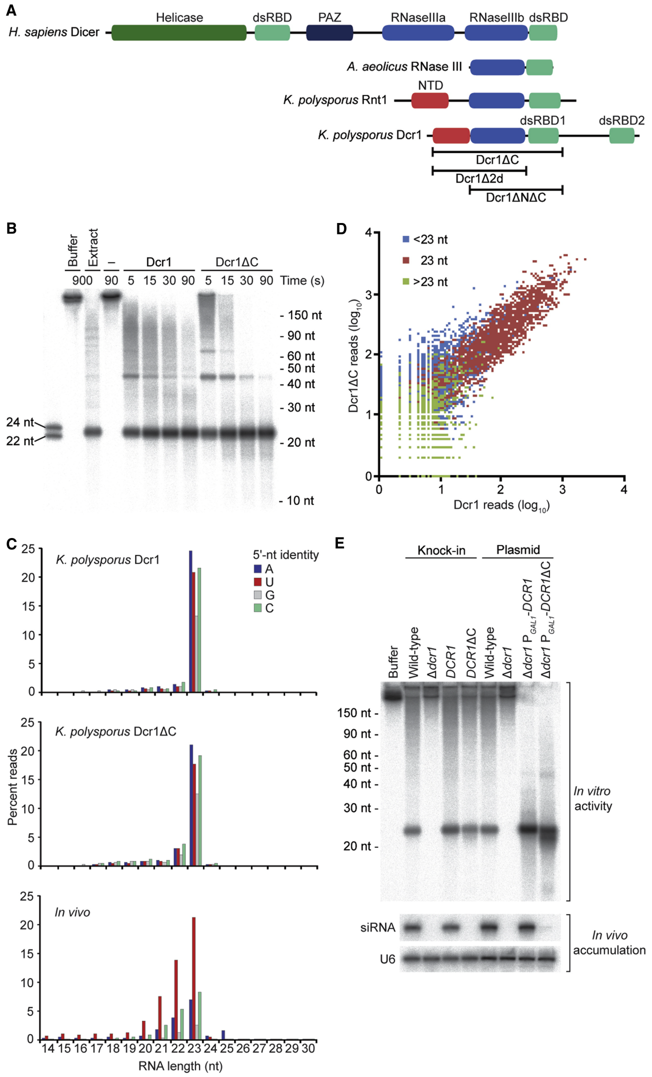


Figure 1. Activity of *K. polysporus* Dcr1ΔC

(A) Domain architectures of representative RNase III proteins. The N-terminal domain (NTD) unique to budding-yeast RNase III enzymes is indicated.

(B) Activity of recombinant Dcr1 proteins under standard single-turnover conditions. Body-labeled 500 bp dsRNA was incubated with no enzyme (–), full-length *K. polysporus* Dcr1, or *K. polysporus* Dcr1ΔC for the indicated time. For comparison, the substrate was incubated with *K. polysporus* whole-cell extract (Extract) or with the buffer used for extracts (Buffer). Also shown are radiolabeled synthetic 22 and 24 nt RNAs (left) and the migration of other RNA standards (right).

(C) Length distributions of products with the indicated 5′ nucleotides. Top and middle: Substrate-matching sequencing reads from analysis of in vitro reactions using the indicated purified proteins. Bottom: For reference, genome-matching small RNAs sequenced from *K. polysporus*, redrawn from Drinnenberg et al. (2009).

(D) Correlation between cleavage products generated by purified *K. polysporus* Dcr1 and Dcr1ΔC. Plotted is the read count (including a pseudocount of 1) for each sequenced product from the Dcr1ΔC reaction (y axis) compared to the count for the corresponding product from the

Dcr1 reaction (x axis). Colors indicate the length of the sequenced products: 14–22 nt (blue), 23 nt (red), and 24–30 nt (green).

(E) Comparison of in vitro activity with product accumulation in strains expressing the corresponding Dcr1 variants. Top: Body-labeled 500 bp dsRNA was incubated with extracts from *S. castellii* strains with the indicated deletions, additions, and replacements. Bottom: An RNA blot with samples from the same strains was probed for an endogenous siRNA, then reprobed for U6 small nuclear RNA.

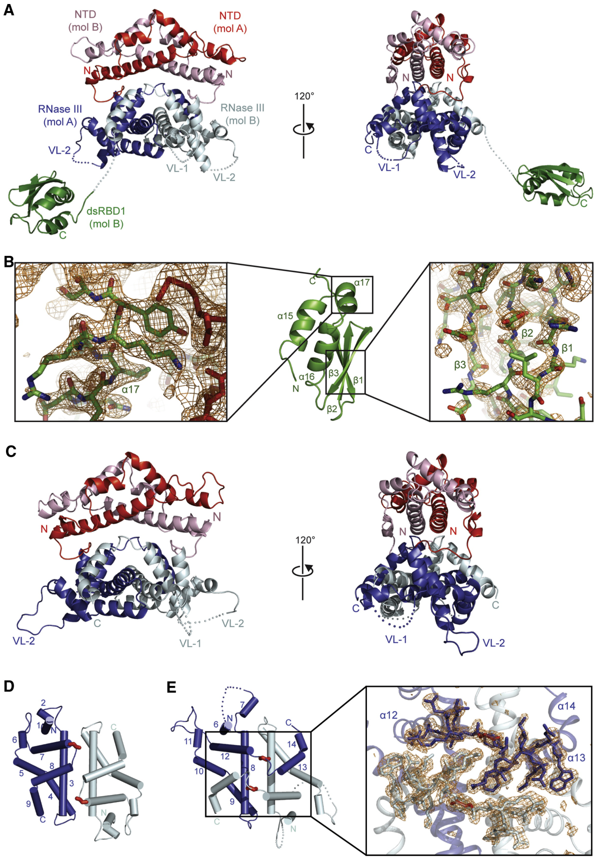


Figure 2. Crystal structures of Dcr1ΔC and E224Q Dcr1Δ2d

(A) Crystal structure of a Dcr1ΔC dimer at 2.3 Å resolution, showing a pair of NTDs (red and pink), a pair of RNase III domains (blue and silver), and a single dsRBD1 (green). The other dsRBD1 had poor density and is not shown. Disordered loops are shown as dotted lines, and VL-1 and VL-2 are labeled.

(B) Structure of Dcr1ΔC dsRBD (middle), flanked by close-up views of representative segments with good (left) or poor (right) electron density maps (2*F*o–*F*c). This domain is stabilized through interactions with a symmetry-related NTD of another Dcr1ΔC molecule (red sticks).

(C) Crystal structure of an E224Q Dcr1Δ2d dimer at 1.97 Å resolution. Domain designations and colors are as in (A).

(D) Topology of the *A. aeolicus* RNase III domain dimer (PDB code 2NUG). Individual monomers are colored in blue and silver, with glycine residues at the kink point between helices α7 and α8 highlighted (red stick models).

(E) Topology of *K. polysporus* E224Q Dcr1Δ2d. Helices α13 and α14 (numbered according to the entire model) participate in a domain swap, with kink-point glycine residues indicated as in (D). The expanded region shows an *F*o–*F*c simulated annealing omit map of the loops between helices α12 and α13 contoured at 3.2σ (orange wire mesh). The α12-α13 loops are shown as stick models, and the other regions are shown as ribbon models.

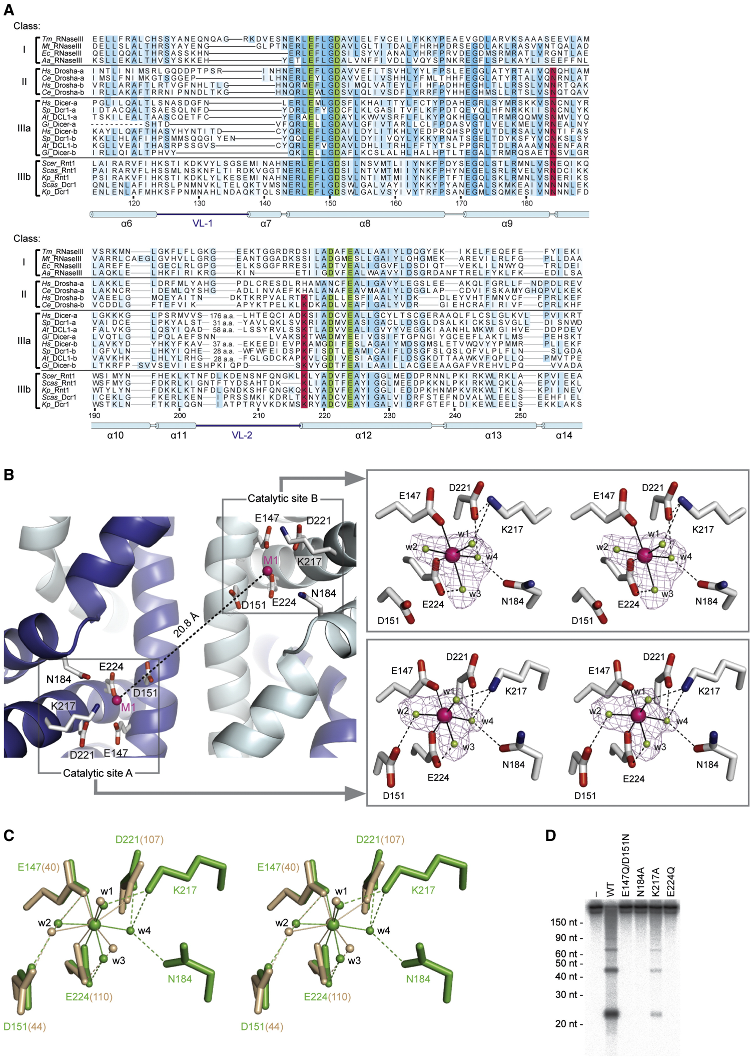


Figure 3. The active site of *K. polysporus* Dcr1

(A) Sequence alignment of RNase III domains from each class of RNase III enzymes. Highlighted are the previously identified catalytic residues (green), two newly identified catalytic residues (red), and additional well-conserved amino acids (blue; intensity indicates degree of conservation). Residue numbers and the secondary structure of *K. polysporus* Dcr1 are indicated below the alignment. Tandem RNase III domains present in Drosha and canonical Dicer are designated a and b. *Tm*, *Thermotoga maritoma*; *Mt*, *Mycoplasma tuberculosis*; *Ec*, *Escherichia coli*; *Aa*, *Aquifex aeolicus*; *Hs*, *Homo sapiens*; *Ce*, *Caenorhabditis elegans*; *Sp*, *Schizosaccharomyces pombe*; *At*, *Arabidopsis thaliana*; *Gi*, *Giardia intestinalis*; *Scer*, *Saccharomyces cerevisiae*; *Scas*, *S. castellii*; and *Kp*, *K. polysporus*.

(B) Left: Close-up view of the Dcr1 catalytic sites showing RNase III domains (ribbons), conserved catalytic residues (sticks), and Mg^{2+} ions (spheres). Right: Metal-ion coordination in the active sites shown in stereo view. Octahedral coordination bonds and hydrogen bonds are drawn as solid and dotted lines, respectively. *F*o–*F*c simulated annealing omit maps of Mg^{2+} metal ion (magenta) and four water molecules (lime) are contoured at 4.0σ (purple wire mesh).

(C) Stereo view of the superimposed active sites of apo-Dcr1ΔC (green) and apo-*Aa*RNase III (tan; PDB code 1RC5). Bonds are drawn as in (B).

(D) Dicing activity of active-site mutants under standard single-turnover conditions. Body-labeled 500 bp dsRNA was incubated for 5 s with buffer only (–), wild-type Dcr1ΔC (WT), or Dcr1ΔC variants with the indicated mutations.

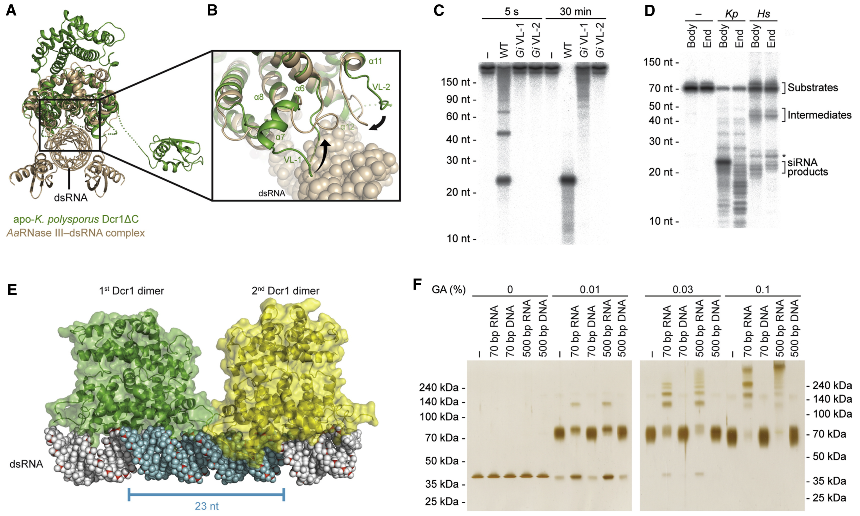


Figure 4. Binding and cleavage of a dsRNA duplex

(A) Structural alignment of apo-Dcr1ΔC (green) and *Aa*RNase III-dsRNA complex (tan) based on their RNase III domains.

(B) Close-up view of the VL-1 and VL-2 loops. The dsRNA is represented with a sphere model. The numbering of secondary structure elements is as in Figure 3A.

(C) Dicing activity of variable-loop mutants under standard single-turnover conditions. Body-labeled 500 bp dsRNA was incubated for 5 s or 30 min with buffer only (–), wild-type Dcr1ΔC (WT), or Dcr1ΔC variants with the indicated *Gi*Dicer loop substitutions.

(D) Cleavage reactions following body- or end-labeled 70 bp dsRNA substrates. Substrates were incubated with buffer only (–), *K. polysporus* Dcr1ΔC (*Kp*; standard single-turnover conditions), or *H. sapiens* Dicer (*Hs*). *Hs*Dicer generates ~21–22 nt siRNA products and ~40–50 nt intermediates as indicated; an ~25 nt product of unknown origin is also observed (*).

(E) A pair of Dcr1ΔC dimers (green and yellow) modeled with dsRNA (sphere model). The anticipated 23 nt siRNA product is highlighted (cyan). The dsRBD is not shown for clarity.

(F) Protein crosslinking analyses of Dcr1ΔC oligomerization. E224Q Dcr1ΔC was incubated with buffer only (–) or the indicated nucleic acid before crosslinking with the indicated glutaraldehyde (GA) concentration. Reactions were analyzed by SDS-PAGE and visualized by silver staining. Also shown is the migration of protein standards with the indicated molecular weights.

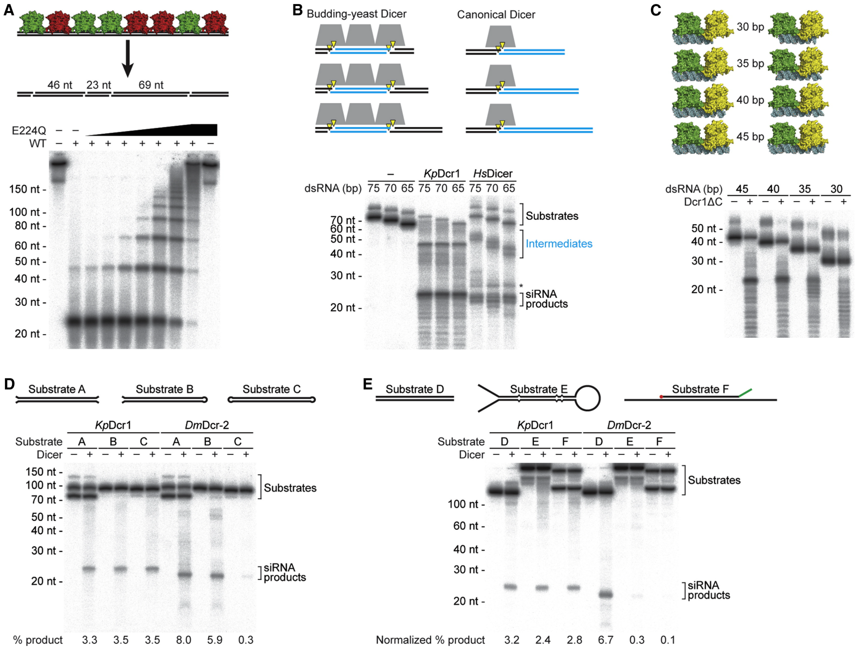


Figure 5. The relationship between active-site spacing and product length

(A) Effects of changing active-site spacing. Top: The schematic depicts wild-type (green) and active-site mutant (red) Dcr1ΔC dimers assembled on dsRNA to generate different products with lengths varying by 23 nt increments. Bottom: Body-labeled 500 bp dsRNA present at 140 pM was incubated without (–) or with (+) 30 nM wild-type Dcr1ΔC in the absence (–) or presence of increasing concentrations of E224Q Dcr1ΔC (3-fold dilution series from 0.33 nM to 243 nM). Products were resolved by denaturing 10% PAGE.

(B) Length of dicing intermediates. Top: The schematic shows the predicted length variation of intermediates (cyan) for substrates of increasing length. Trapezoids represent dimers of RNase III domains. Bottom: Body-labeled dsRNA substrates of the indicated length (present at ~45 pg/μl) were incubated with buffer only (–), a mixture of 30 nM wild-type Dcr1ΔC and 3 nM E224Q Dcr1ΔC (*Kp*Dcr1), or *H. sapiens* Dicer (*Hs*Dicer).

(C) Substrate length requirements for preferentially generating 23 nt products. Top: Docking models depict a pair of dimers bound to the indicated substrates, either anchoring the first dimer at the dsRNA terminus (left) or centering the dimer pair (right). Bottom: Body-labeled dsRNA substrates of the indicated length were incubated for 2 min either without (–) or with (+) Dcr1ΔC under standard single-turnover conditions.

(D) Dicer activities on open and closed substrates. Top: Schematic of substrates, which contained 70 bp of dsRNA flanked by either short ssRNA overhangs or loops. Bottom: Body-labeled substrates were incubated without (–) or with (+) Dcr1ΔC (*Kp*Dcr1) or *D. melanogaster* Dcr-2 (*Dm*Dcr-2) under multiple-turnover conditions (30 nM substrate and 10 nM protein). Size markers were estimated based on RNA standards in Figure S5C.

(E) Dicer activities on substrates resembling endogenous yeast substrates. Top: Schematic of substrates, which contained 161 bp of dsRNA within either a perfect duplex (substrate D), a palindromic RNA (substrate E), or an internal duplex (substrate F; red, 7-methylguanosine cap; green, poly(A) tail). Bottom: Body-labeled substrates were reacted with Dicer enzymes as in (D). Percent product was normalized to account for radiolabeled phosphodiester linkages occurring outside of the dsRNA region of substrate E.

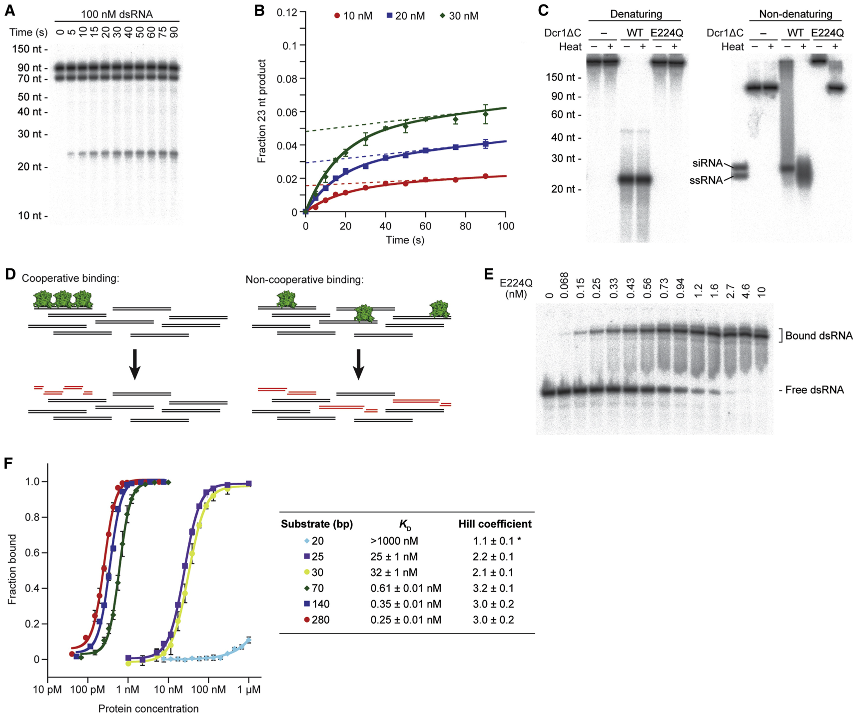


Figure 6. Mechanisms for preventing off-pathway substrate cleavage

(A) Product accumulation under multiple-turnover conditions. Body-labeled 70 bp dsRNA present at 100 nM (implying an ~300 nM concentration of nonoverlapping 23 bp binding sites) was incubated with 30 nM Dcr1ΔC for the indicated time.

(B) Quantitative analysis of multiple-turnover kinetics. Body-labeled 70 bp dsRNA present at 100 nM was incubated with the indicated concentration of Dcr1ΔC for the indicated time. Plotted are average values (n = 3; error bars represent the standard deviation). Solid lines, least-squares fit to the burst equation. Dotted lines, extrapolation of the steady-state rate to the *y* axis.

(C) siRNA products bound to Dcr1ΔC. Body-labeled 500 bp dsRNA was incubated with buffer only (–), wild-type Dcr1ΔC (WT), or an active-site mutant of Dcr1ΔC (E224Q) under standard single-turnover conditions. Reactions were placed on ice (–) or heat-denatured at 90°C for 2 min (+) before fractionation. Products were resolved by denaturing 15% PAGE (left) or native 6% PAGE (right). siRNA, duplex of 23 nt siRNAs; ssRNA, 23 nt ssRNA.

(D) Schematic illustrating that for an enzyme that does not recognize the end of a duplex, siRNA generation under multiple-turnover conditions implies binding cooperativity.

(E) Representative gel-shift analysis of E224Q Dcr1ΔC binding to 70 bp dsRNA. Trace amounts of body-labeled 70 bp dsRNA were incubated with increasing amounts of protein.

(F) Binding isotherms for E224Q Dcr1ΔC. Data points represent average values (n = 3 for all substrates except 30 bp, for which n = 2; error bars represent the standard deviation). Solid lines show the best fit to the Hill equation, which produced the parameters shown to the right. For the 20 bp substrate, the maximum fraction bound was assumed to be 1.0 in order to obtain an estimate of the Hill coefficient (*).

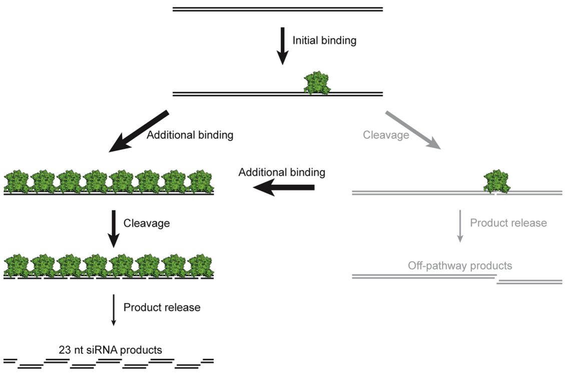


Figure 7. Inside-out mechanism of budding-yeast Dicers
See the main text for a description. Off-pathway cleavage products (gray) are disfavored by both cooperative binding and slow product release.

References

Akey, D.L., and Berger, J.M. (2005). Structure of the nuclease domain of ribonuclease III from M. tuberculosis at 2.1 A. Protein Sci. *14*, 2744–2750.

Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., and Ji, X. (2001). Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure *9*, 1225–1236.

Blaszczyk, J., Gan, J., Tropea, J.E., Court, D.L., Waugh, D.S., and Ji, X. (2004). Noncatalytic assembly of ribonuclease III with double-stranded RNA. Structure *12*, 457–466.

Carmell, M.A., and Hannon, G.J. (2004). RNase III enzymes and the initiation of gene silencing. Nat. Struct. Mol. Biol. *11*, 214–218.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. Mol. Cell *42*, 172–184.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798–802.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544–550.

Du, Z., Lee, J.K., Tjhen, R., Stroud, R.M., and James, T.L. (2008). Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. Proc. Natl. Acad. Sci. USA *105*, 2391–2396.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818–822.

Gan, J., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S., and Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. Cell *124*, 355–366.

Gan, J., Shaw, G., Tropea, J.E., Waugh, D.S., Court, D.L., and Ji, X. (2008). A stepwise model for double-stranded RNA processing by ribonuclease III. Mol. Microbiol. *67*, 143–154.

Halic, M., and Moazed, D. (2010). Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell *140*, 504–516.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

Hopwood, D. (1975). The reactions of glutaraldehyde with nucleic acids. Histochem. J. *7*, 267–276.

Lamontagne, B., Tremblay, A., and Abou Elela, S. (2000). The N-terminal domain that distinguishes yeast from bacterial RNase III contains a dimerization signal required for efficient double-stranded RNA cleavage. Mol. Cell. Biol. *20*, 1104–1115.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69–81.

Leulliot, N., Quevillon-Cheruel, S., Graille, M., van Tilbeurgh, H., Leeper, T.C., Godin, K.S., Edwards, T.E., Sigurdsson, S.T., Rozenkrants, N., Nagel, R.J., et al. (2004). A new a-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III. EMBO J. *23*, 2468–2477.

MacRae, I.J., and Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr. Opin. Struct. Biol. *17*, 138–145.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195–198.

MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. Nat. Struct. Mol. Biol. *14*, 934–940.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656–668.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343–349.

Meng, W., and Nicholson, A.W. (2008). Heterodimer-based analysis of subunit and domain contributions to double-stranded RNA processing by Escherichia coli RNase III in vitro. Biochem. J. *410*, 39–48.

Nussbaum, A.K., Dick, T.P., Keilholz, W., Schirle, M., Stevanovi_c, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D.H., Huber, R., et al. (1998). Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. Proc. Natl. Acad. Sci. USA *95*, 12504–12509.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803–806.

Sun, W., and Nicholson, A.W. (2001). Mechanism of action of Escherichia coli ribonuclease III. Stringent chemical requirement for the glutamic acid 117 side chain and Mn2+ rescue of the Glu117Asp mutant. Biochemistry *40*, 5102–5110.

Sun, W., Li, G., and Nicholson, A.W. (2004). Mutational analysis of the nuclease domain of Escherichia coli ribonuclease III. Identification of conserved acidic residues that are important for catalytic function in vitro. Biochemistry *43*, 13054–13062.

Takeshita, D., Zenno, S., Lee, W.C., Nagata, K., Saigo, K., and Tanokura, M. (2007). Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. J. Mol. Biol. *374*, 106–120.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529.

Vaucheret, H. (2005). MicroRNA-dependent trans-acting siRNA production. Sci. STKE *2005*, pe43.

Wang, H.W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009a). Structural insights into RNA processing by the human RISC-loading complex. Nat. Struct. Mol. Biol. *16*, 1148–1153.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009b). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754–761.

Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. Mol. Cell *41*, 589–599.

Wenzel, T., Eckerskorn, C., Lottspeich, F., and Baumeister, W. (1994). Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. FEBS Lett. *349*, 205–209.

Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004). Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. Proc. Natl. Acad. Sci. USA *101*, 8307–8312.

Xiao, J., Feehery, C.E., Tzertzinis, G., and Maina, C.V. (2009). E. coli RNase III(E38A)generates discrete-sized products from long dsRNA. RNA *15*, 984–991.

Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proc. Natl. Acad. Sci. USA *99*, 9942–9947.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. *21*, 5875–5885.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell *118*, 57–68.

Extended Experimental Procedures

Protein purification

A list of expression plasmids generated in this study is provided (Table S4). The gene encoding *K. polysporus* Dcr1ΔC was cloned into a modified pRSFDuet vector (Novagen) containing an amino-terminal Ulp1-cleavable His_6-Sumo tag. Protein was overexpressed in *E. coli* BL21(DE3) Rosetta2 (Novagen). Cells expressing the recombinant protein were lysed by sonication in Buffer A (10 mM phosphate buffer pH 7.3, 640 mM NaCl, 25 mM imidazole, 10 mM β-mercaptoethanol (β-ME), 1 mM phenylmethylsulphonyl fluoride), and the lysates were cleared by centrifugation. The supernatant was loaded onto a nickel column and then washed with Buffer A. The target protein was eluted with a linear gradient of 25 mM to 1 M imidazole. The protein fractions were collected and dialyzed against Buffer B (10 mM phosphate buffer pH 7.3, 500 mM NaCl, 20 mM imidazole, 10 mM β-ME) overnight. Ulp1 protease was added, and the digested protein was loaded onto a nickel column to remove the cleaved His_6-Sumo tag. Ammonium sulfate was added to the flow-through containing Dcr1ΔC, and the sample was centrifuged. The supernatant was loaded onto a phenyl-sepharose hydrophobic interaction column in Buffer C (10 mM phosphate buffer pH 7.3, 2 M ammonium sulfate, 10 mM β-ME), and the protein was eluted with a linear gradient of 1 to 0 M ammonium sulfate. The eluted protein was dialyzed against Buffer D (100 mM Na/K-phosphate buffer pH 7.3, 10 mM β-ME) and then loaded onto a Heparin-affinity column in Buffer B. The protein was eluted with a linear gradient of 0 to 2 M NaCl. The eluted sample was dialyzed against Buffer D and then loaded onto a MonoQ column in Buffer D. The protein was eluted with a linear gradient of 0 to 2 M NaCl. The eluted sample was concentrated and loaded onto a HiLoad 200 16/60 column in Buffer E (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT). The final two purification steps

(MonoQ and HiLoad 200 columns) were performed 1–3 times, until contaminating nucleic acids were no longer present as judged by UV absorbance (A_{280}/A_{260}). Purified Dcr1ΔC was concentrated to approximately 20 mg ml^{-1} using ultrafiltration and stored at −80°C in Protein Storage Buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT).

Point substitutions were introduced using site-directed mutagenesis. Full-length Dcr1, selenomethionine (SeMet)-substituted Dcr1ΔC, active-site Dcr1ΔC mutants, Dcr1ΔNΔC, and Dcr1ΔC variants containing *Gi*Dicer VL-1 or VL-2 with or without the E224Q mutation were purified essentially as described above for wild-type Dcr1ΔC, except full-length Dcr1 contained contaminating nucleic acids and was stored in a different buffer (10 mM Tris-HCl pH 8.5, 600 mM NaCl, 10 mM β-ME). Both of the Dcr1ΔC VL-1 and VL-2 variants eluted from the gel filtration column at the same retention time as wild-type Dcr1ΔC, suggesting that they are correctly folded (data not shown). Dcr1Δ2d and E224Q Dcr1Δ2d were overexpressed as His_6-Sumo tag fusion proteins as described for Dcr1ΔC. After removal of the cleaved His_6-Sumo tag, the proteins were loaded onto a Heparin-affinity column in Buffer D. The protein was eluted with a linear gradient of 0 to 2 M NaCl. Ammonium sulfate was added to the eluted fractions, and the sample was centrifuged. The supernatants were loaded onto a phenyl-sepharose hydrophobic interaction column in Buffer C, and the protein was eluted with a linear gradient of 1 to 0 M ammonium sulfate. The eluted samples were concentrated and loaded onto a HiLoad 200 16/60 column in Buffer E.

Structure determination and refinement

Native crystals of Dcr1ΔC were obtained at 20°C by sitting-drop vapour diffusion in 0.18 M triammonium citrate, 20% PEG3350, 10 mM Na-HEPES buffer (pH 7.5), 20 mM magnesium chloride, and 3% 2-propanol. SeMet-substituted crystals were grown at 20°C by sitting-drop vapour diffusion in 220 mM Na-malonate buffer (pH 7.0), 15% PEG4000, and 10 mM phenol. The native and SeMet-substituted crystals of Dcr1ΔC were soaked in collection buffer (1.2-fold concentrated reservoir solution), cryoprotected with 10% glycerol, and flash-cooled in a nitrogen stream at 100 K. Both derivative data sets were collected at Brookhaven NSLS beamline X29. Data were processed with the program HKL2000 (Otwinowski and Minor, 1997). Data collection and refinement statistics are shown in Table S2. A total of 20 selenium sites were found using peak data with the program SnB (Weeks and Miller, 1999) and were used for phase calculation at 3.5 Å resolution with the program SHARP (de La Fortelle and Bricogne, 1997). The initial phases were improved by solvent flattening with the program SOLOMON (Abrahams and Leslie, 1996) and by non-crystallographic symmetry averaging with the program CCP4 (Collaborative Computational Project, 1994). The initial model was built manually with the programs O (Jones et al., 1991) and Coot (Emsley and Cowtan, 2004) and was improved by iterative cycles of refinement with the program Phenix (Adams et al., 2002), which provided a clear electron density map for the N-terminal and RNase III domains. Molecular replacement was performed with the program MOLREP (Vagin and Teplyakov, 2000), using the SeMet structure as a search model. Further model building and refinement revealed electron peaks in the F_o–F_c density map for one of the four dsRBD1s in the asymmetric unit. The model using the native data was improved to 2.3 Å resolution for the four Dcr1ΔC molecules in the asymmetric unit. Two of the four molecules had completely disordered dsRBD1s, one had a poorly defined dsRBD1, and the other had a well-defined dsRBD1; only the latter was included in the final coordinates. The Ramachandran plot analysis by PROCHECK (Collaborative Computational Project, 1994) showed 93.1% and 6.9% of the protein residues in the most favourable and additionally allowed regions, respectively, with no residues in disallowed regions. The *F*o−*F*c simulated annealing omit map was calculated by CNS (Brunger et al., 1998), and the resultant map showed clear octahedral coordination around active-site residues.

Native crystals of E224Q Dcr1Δ2d were obtained at 20°C by sitting-drop vapour diffusion in 0.2 M L-proline, 0.1 M HEPES pH 7.5, and 10% PEG3350. The crystals were soaked in collection buffer (1.2-fold concentrated reservoir solution), cryoprotected with 30% glycerol, and flash-cooled in a nitrogen stream at 100 K. The derivative data set was collected at Advanced Photon Source NE-CAT beamlines. Molecular replacement performed with MOLREP (Vagin and Teplyakov, 2000) using the structure of the NTD and RNase III domains from Dcr1ΔC as a search model revealed the entire molecule. Model building and refinement were done using Coot (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2002), respectively. The electron density for the swapped segment within the RNase III domain dimer could be clearly traced in the simulated annealing omit map (*F*o−*F*c) with CNS (Brunger et al., 1998).

dsRNA modeling

In Figure 4E, two molecules of the *Aa*RNase III–dsRNA complex (PDB code 2NUF) and a 47 bp dsRNA (PDB code 3CIY) (Liu et al., 2008) were structurally aligned based on their bound dsRNA such that the RNA cleavage sites were 23 nt apart on the dsRNA. Then, two Dcr1ΔC dimers were structurally aligned onto the homodimeric RNase III domains of the docked *Aa*RNase III molecules.

Preparation of RNA substrates

dsRNA substrates were prepared by annealing of ssRNA generated by *in vitro* transcription with T7 RNA polymerase. PCR-generated templates were used for transcription of >76 nt RNAs, and gel-purified DNA oligonucleotides were used for transcription of shorter RNAs. Transcription reactions for substrates used in single-turnover processing reactions were assembled using the MAXIscript Kit (Ambion) with a 61:1 molar ratio of UTP:[α-^{32}P]UTP (800 Ci/mmol) according to the manufacturer's directions. High specific-activity substrates for use in binding reactions were transcribed with a 2:1 (70–280 nt RNAs) or 16:1 (20–30 nt RNAs) molar ratio of UTP:[α-^{32}P]UTP (800 Ci/mmol). Trace-labeled RNA used for analysis of multiple-turnover kinetics was transcribed with a 3125:1 molar ratio of UTP:[α-^{32}P]UTP. DNase-treated ssRNA was fractionated by denaturing PAGE, eluted from gel slices in 0.3 M NaCl overnight at 4°C, and ethanol precipitated. Complementary RNAs were combined in dsRNA Annealing Buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA), heated to 90°C for 1 min, and slowly cooled to room temperature over 4–5 hr. Annealed RNA was fractionated by native PAGE (with the exception of 500 bp dsRNA, which was fractionated on a 4% urea gel), and dsRNA was eluted from gel slices in 0.3 M NaCl overnight at 4°C, ethanol precipitated, and stored in dsRNA Storage Buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA). RNA was quantified by scintillation counting, and 10X stocks were prepared in dsRNA Storage Buffer supplemented with 1 μM yeast tRNA (Sigma). Body-labeled dsRNA substrates containing 5'-monophosphate termini were prepared similarly, except that gel-purified ssRNA was treated with calf intestinal phosphatase, phosphorylated with T4 polynucleotide kinase, and gel-purified by denaturing PAGE. End-labeled dsRNA substrates were prepared similarly, except radiolabeled UTP was omitted from transcription reactions and phosphorylation was performed with a ~32:1

molar ratio of ATP:[γ-^{32}P]ATP (6000 Ci/mmol) before chasing with 1 mM ATP. Substrates used in glutaraldehyde crosslinking experiments were prepared as described in the section on glutaraldehyde crosslinking.

The substrates used in Figure 5D were initially prepared as a single body-labeled dsRNA substrate (1724:1 molar ratio of UTP:[α-^{32}P]UTP) containing 5'-monophosphate termini as described above. This dsRNA was treated with T4 RNA Ligase 1 (NEB) and products were resolved by denaturing PAGE (8% acrylamide, 90% formamide). The incomplete ligation reaction gave rise to unligated duplex (substrate A), hairpin RNA (substrate B), and closed dsRNA (substrate C), which were distinguished by comparison with control ligation reactions using dsRNA containing 5'-triphosphate termini at one or both ends (Figure S5A). Gel-purified RNA was quantified by scintillation counting, and 300 nM stocks were prepared in dsRNA Storage Buffer. The covalently closed nature of substrate C was subsequently confirmed by limited hydrolysis under alkaline conditions, which nicks the RNA to generate slower-migrating species with mobilities that are comparable to the mobility of the hairpin RNA (Figure S5B).

The substrates used in Figure 5E were designed to contain 161 bp of duplex RNA of equivalent specific activity. Substrate D was prepared as a body-labeled dsRNA (3781:1 molar ratio of UTP:[α-^{32}P]UTP) containing 5'-monophosphate termini as described above. The DNA template used for transcription of substrate E was prepared by PCR amplification of supercontig712:108196–108786(–) from *S. castellii* genomic DNA, with the addition of an upstream T7 promoter. The palindromic RNA substrate was transcribed with a 3946:1 molar ratio of UTP:[α-^{32}P]UTP; DNase-treated RNA was fractionated by denaturing PAGE (8% acrylamide, 90% formamide); and the full-length product was eluted from gel slices in 0.3 M NaCl. Substrate F consisted of a body-labeled capped and polyadenylated (A_{20}) ssRNA annealed to an unlabeled ssRNA. The body-labeled RNA was prepared by transcription with a 1579:1 molar ratio of UTP:[α-^{32}P]UTP from a PCR template coding for 20 terminal A residues. Gel-purified ssRNA was enzymatically capped using the ScriptCap m^{7}G Capping System (Epicentre Biotechnologies) and purified by denaturing PAGE. Unlabeled ssRNA was transcribed without radiolabeled UTP and purified by denaturing PAGE. The labeled strand was combined with a 10% excess of the unlabeled strand in dsRNA Annealing Buffer, dsRNA was annealed by heating to 90°C for 1 min and slowly cooling to room temperature, and RNA was concentrated by ethanol precipitation. All RNAs were resuspended in dsRNA Storage Buffer and quantified by scintillation counting, and 300 nM stocks were prepared in dsRNA Storage Buffer.

Dicer activity assays

Yeast extracts. Preparation of whole-cell extracts and processing reactions were performed as described in (Drinnenberg et al., 2009) with the following modifications. Substrates were used at a final concentration of ~45 pg/µl (corresponding to ~140 pM for 500 bp dsRNA and ~280 pM for 500 nt ssRNA). Reactions were incubated at room temperature (22–24°C) for 15 min (*K. polysporus*) or 2 hr (*S. castellii*) and stopped by addition of phenol-chloroform and EDTA.

K. polyporus *Dcr1*. For biochemical assays, proteins were diluted and stored at –20°C in Protein Dilution Buffer (5 mM Tris-HCl pH 7.5, 100 mM NaCl, 2.5 mM DTT, 50% glycerol, 1 mg/ml Ultrapure BSA [Ambion]). All Dcr1 protein concentrations are expressed in terms of dimer concentration. 10 µl reactions contained 2 µl 5X reaction buffer (150 mM Tris-HCl pH 7.5, 150 mM NaCl, 25 mM $MgCl_2$, 5 mM DTT, 0.5 mM EDTA), 1 µl Dcr1 protein or Protein Dilution Buffer, and 1 µl RNA substrate. All reactions were under standard single-turnover conditions, unless otherwise indicated. Standard single-turnover reactions contained protein at a final concentration of 30 nM and RNA at a final concentration of ~45 pg/µl (corresponding to ~1 nM for 70 bp dsRNA). Reactions were incubated at room temperature (22–24°C) for 90 s unless otherwise indicated and quenched by addition to 10 µl Formamide Loading Buffer (90% formamide, 18 mM EDTA, 0.025% sodium dodecyl sulfate, 0.1% xylene cyanol, 0.1% bromophenol blue). Multiple-turnover reactions in Figures 6A–B contained protein at a final concentration of 10–30 nM as indicated and 70 bp dsRNA at a final concentration of ~4.5 ng/µl (corresponding to ~100 nM); multiple-turnover reactions in Figure S6E were supplemented with NaCl to obtain the indicated final concentration, and contained protein at a final concentration of 30 nM and 500 bp dsRNA at a final concentration of ~4.5 ng/µl (corresponding to ~14 nM). Reactions were incubated at room temperature, and 4 µl aliquots were removed at the indicated time and quenched by addition to 12 µl Formamide Loading Buffer. Multiple-turnover reactions in Figures 5D, 5E, and S5C contained protein at a final concentration of 10 nM and substrates at a final concentration of ~30 nM. 5 µl reactions were incubated at room temperature for 90 s and quenched by addition of 15 µl Formamide Loading Buffer.

Human Dicer. Recombinant *H. sapiens* Dicer (0.5 U/µl, Genlantis) was combined with an equal volume of Protein Dilution Buffer. 10 µl reactions contained 2 µl 5X reaction buffer (150 mM Tris-HCl pH 6.8, 125 mM NaCl, 15 mM $MgCl_2$, 5 mM DTT, 0.5 mM EDTA, 5% glycerol), 1 µl Dicer (0.25 U/µl), and 1 µl RNA substrate. Reactions were incubated at 37°C for 15 min (Figure 5B) or 30 min (Figure 4D) and quenched by addition to 10 µl Formamide Loading Buffer.

Drosophila *Dcr-2*. Purified recombinant *D. melanogaster* Dcr-2 (expressed using the baculovirus system) was a generous gift from Phil Zamore (University of Massachusetts Medical

Center). 5 µl reactions contained 1 µl 5X reaction buffer (150 mM Tris-HCl pH 6.8, 125 mM NaCl, 15 mM $MgCl_2$, 5 mM DTT, 0.5 mM EDTA, 5% glycerol), 0.5 µl 10 mM ATP, 0.5 µl *Dm*Dcr-2 protein or Dcr-2 Storage Buffer (60 mM HEPES-KOH pH 7.4, 400 mM NaCl, 4 mM DTT, 50% glycerol), and 0.5 µl RNA substrate. *Dm*Dcr-2 was present at a final concentration of ~10.6 nM, and RNA was at a final concentration of ~30 nM for each substrate. Reactions were incubated at room temperature for 30 min (Figure 5D and Figure S5C) or 5 min (Figure 5E) and quenched by addition of 15 µl Formamide Loading Buffer.

Analysis. RNA products were resolved by 7.5 M urea 15% PAGE, unless otherwise indicated. Radiolabeled products were visualized by phosphorimaging (Fujifilm BAS-2500) and quantified using Multi Gauge (Fujifilm). To quantify product formation in Figures 5D, 5E, and S5C, the percent product was measured as P_P = 100*[siRNA product]/([siRNA product] + [substrate]), and the background P_P value (calculated from the corresponding –Dicer control lane) was subtracted. For quantitative analysis of multiple-turnover kinetics in Figure 6B, at each time point (t) the fraction product was measured as F_P = [23 nt product]/([23 nt product] + [70 nt substrate]). Data were fit to the burst equation:

$$F_P = A(1 - e^{-kt}) + bt$$

where A was the burst amplitude, k was the exponential burst rate constant, and b was the linear steady-state velocity. Fits were performed using the nonlinear least-squares method implemented in KaleidaGraph (Synergy Software).

Dicer binding assays

10 µl reactions contained 2 µl 5X binding buffer (150 mM Tris-HCl pH 7.5, 150 mM NaCl, 25 mM $MgCl_2$, 5 mM DTT, 0.5 mM EDTA, 25% glycerol), 1 µl protein, and 1 µl RNA

substrate, with the exception of reactions in Figure S6D which were supplemented with an additional 109 mM NaCl. E224Q Dcr1ΔC was present at the indicated final concentration, and RNA was present at a final concentration of ~450 fg/µl (corresponding to ~10 pM for 70 bp dsRNA) for 70–280 bp substrates or ~11.6 pg/ul (corresponding to ~400 pM for 30 bp dsRNA) for 20–30 bp substrates. Reactions were incubated at room temperature (22–24°C) for 10 min and then on ice for at least an additional 20 min. Reactions were analyzed on native polyacrylamide gels (6% 29:1 acrylamide:bis-acrylamide, 0.5X TBE, 5 mM $MgCl_2$, 5% glycerol) cooled to 4°C. Samples were loaded without addition of loading buffer, and gels were run at 10W for 2.5 hr at 4°C in 0.5X TBE supplemented with 5 mM $MgCl_2$. RNA was visualized by phosphorimaging (Fujifilm BAS-2500) and quantified using Multi Gauge (Fujifilm). At each concentration (P), the fraction bound (as designated in Figure S6B) was measured as $F_B = \text{dsRNA}_{\text{bound}}/(\text{dsRNA}_{\text{bound}} + \text{dsRNA}_{\text{free}})$. Data were fit to the Hill equation:

$$F_B = A + B/[1 + (K_D/P)^n]$$

where A was background signal, B was the magnitude of the dynamic range, K_D was the apparent dissociation constant, and n was the Hill coefficient. To fit binding data for the 20 bp dsRNA substrate, B was set equal to 1.0. Fits were performed using the nonlinear least-squares method implemented in KaleidaGraph (Synergy Software).

Multi-angle light scattering

Molecular-weight experiments were undertaken on a three-angle light scattering detector (mini-DAWN EOS) and refractive index detector (Optilab DSP, Wyatt Technology). Data were collected every 0.5 s at a flow rate of 0.2 mL per min. Data analysis was carried out using the program ASTRA, yielding the molar mass and mass distribution of the sample. The middle portion of the peak saturated the refractive-index detector, and hence the analysis was restricted to the shoulders of the peak (Figure S2B). The red and green lines represent molecular mass as a function of elution volume.

Glutaraldehyde crosslinking

Radiolabeled Nucleic Acids. dsRNA substrates were prepared by annealing of body-labeled ssRNA containing 5'-monophosphate termini, which were generated as described above except that 1257:1 and 1571:1 molar ratios of UTP:[α-^{32}P]UTP were used for transcription of 70 nt and 500 nt ssRNAs, respectively. 70 bp dsDNA was prepared by annealing of end-labeled ssDNA, which was generated by phosphorylating gel-purified DNA oligonucleotides using a 571:1 molar ratio of ATP:[γ-^{32}P]ATP (6000 Ci/mmol) before chasing with 1 mM ATP. End-labeled 500 bp dsDNA was prepared by phosphorylating PCR primers using a 77:1 molar ratio of ATP:[γ-^{32}P]ATP (6000 Ci/mmol) before chasing with 1 mM ATP, and then performing PCR using the end-labeled primers. Gel-purified dsRNA and dsDNA were eluted from gel slices in 0.3 M NaCl overnight at room temperature, ethanol precipitated, and stored in Crosslinking Storage Buffer (10 mM HEPES-NaOH pH 7.6, 10 mM NaCl, 0.1 mM EDTA). dsDNA and dsRNA were quantified by absorbance at 260 nm, and 10X stocks (986 nM for 70 bp substrates and 138 nM for 500 bp substrates, each corresponding to 3 µM non-overlapping 23 bp fragments) were prepared in Crosslinking Storage Buffer.

Protein-only Reactions. Dcr1ΔC and Dcr1ΔNΔC were prepared as 10 µM stocks in BSA-Free Protein Dilution Buffer (5 mM Tris-HCl pH 7.5, 100 mM NaCl, 2.5 mM DTT, 50% glycerol), stored at –20°C, and centrifuged at 4°C for 10 min at 16100*g* to remove any aggregates before using. 10 µl crosslinking reactions contained 2 µl 5X Crosslinking Reaction

Buffer (150 mM HEPES-NaOH pH 7.6, 150 mM NaCl, 25 mM $MgCl_2$, 5 mM DTT, 0.5 mM EDTA), 0.3 µl 10 µM protein, and 2 µl freshly diluted glutaraldehyde (5X indicated final concentration, Sigma). Reactions were incubated at room temperature for 10 min, quenched by addition of 10 µl Crosslinking Quench Buffer (2X Laemmli Sample Buffer, supplemented with 5% β-ME and 100 mM Tris-HCl pH 8.0), and incubated for an additional 10 min. Quenched reactions were separated by SDS-PAGE (4–12% NuPAGE Novex Bis-Tris Mini Gels in MOPS SDS Running Buffer, Invitrogen) and analyzed by silver staining.

Reactions with RNA/DNA. E224Q Dcr1ΔC was prepared for crosslinking as described above. 44 µl binding reactions contained 11 µl 5X Crosslinking Reaction Buffer, 1.65 µl 10 µM E224Q Dcr1ΔC, and 5.5 µl Crosslinking Storage Buffer or 10X RNA/DNA. Binding reactions were incubated at room temperature for 10–20 min before initiating the crosslinking reactions. 8 µl aliquots of the binding reactions were mixed with 2 µl glutaraldehyde, incubated at room temperature for 10 min, quenched by addition of 10 µl Crosslinking Quench Buffer, and incubated for an additional 10 min. Quenched reactions were separated by SDS-PAGE and analyzed by silver staining. Radiolabeled nucleic acids were visualized by phosphorimaging of the silver-stained gel. To confirm the formation of Dicer-dsRNA complexes, 8 µl aliquots of the binding reactions (including negative control reactions with BSA substituted for E224Q Dcr1ΔC) were mixed with 2 µl 25% glycerol, incubated on ice for 10–20 min, and analyzed on native polyacrylamide gels as described for the dicer binding assays.

Yeast manipulations

Growth conditions and genetic manipulations. S. castellii and *K. polysporus* were grown at 25°C (with the exception of strains in Figure S5D, which were grown at 30°C) on standard *S.*

cerevisiae plate and liquid media (e.g., YPD and SC). Transformations of *S. castellii* were performed as described in (Drinnenberg et al., 2009) with some modifications. Either 1.5 µg plasmid DNA or 10 µg linear DNA was added to 5 µl single-stranded DNA (10 mg/ml salmon sperm DNA, Sigma D7656), mixed with 50 µl yeast (~3 x 10^8 cells in 100 mM lithium acetate), and added to transformation buffer (a mixture of 240 µl 40% PEG 3350 and 36 µl 1 M lithium acetate). After incubation at 25°C for 30 min, 35 µl of DMSO was added, and the entire mixture was incubated at 30°C for 20 min, resuspended, and then plated on selective media.

Strain construction. A list of strains generated or used in this study is provided (Table S3). A haploid strain expressing *S. castellii* Dcr1ΔC (comprising Tyr17–Glu356) from its native promoter was constructed by two-step homologous recombination in DPB277, as follows: An *S. cerevisiae URA3* expression cassette (amplified from pYES2.1, Invitrogen) was used to replace the open reading frame of *DCR1* by transformation and selection of transformants on SC–ura plates to generate DPB278; the *URA3* cassette was subsequently replaced by the open reading frame encoding Dcr1ΔC by transformation and selection on 5-FOA to generate DPB437. A control strain in which the full-length open reading frame of *DCR1* was used for replacement (DPB406) was generated similarly.

Expression plasmid construction. *S. castellii DCR1* or *DCR1ΔC* was cloned into pYES2.1 (Invitrogen) to produce the galactose-inducible expression plasmids pYES2.1-ScDcr1 and pYES2.1-ScDcr1ΔC, respectively. Plasmids were transformed into DPB318, and expression was induced by growth in SC–ura containing 1% galactose and 1% raffinose.

Blots

Strains. Strains used in Figure 1E were DPB277, DPB278, DPB406, DPB437, DPB005, DPB318, DPB318 transformed with pYES2.1-Dcr1, and DPB318 transformed with pYES2.1-Dcr1ΔC; in Figure S1F were DPB005, DPB318, DPB318 transformed with pYES2.1-Dcr1, and DPB318 transformed with pYES2.1-Dcr1ΔN; and in Figure S5D were DPB005, DPB318, DPB318 transformed with pYES2.1-Dcr1, and DPB318 transformed with pAG416Gal-Dicer (Tables S3 and S4).

RNA blots. Total RNA was isolated using the hot-phenol method. Small-RNA blots were performed using 10–20 µg total RNA per lane and carbodiimide-mediated cross-linking to the membrane (Pall et al., 2007). Blots were hybridized with the following DNA probes radiolabeled at their 5' termini: *S. castellii* siRNA sc1056, 5'-CTATCTTCATCGATTACCATCTA; *S. castellii* U6 small nuclear RNA, 5'-TATGCAGGGGAACTGCTGAT. To detect any siRNAs (including siRNA sc1056) deriving from the 5' arm of the palindrome of substrate E (Figure 5E), the RNA blot in Figure S5D was hybridized with a body-labeled RNA probe corresponding to sequence from *S. castellii* sc1056:165468–165756(–), which was complementary to this arm.

Immunoblots. Three OD600 units of cells were resuspended in 100 ml H_2O. After adding 160 µl of extraction buffer (1.85 M NaOH, 7.4% β-mercaptoethanol), cells were incubated on ice for 10 min. 160 µl of 50% trichloroacetic acid was added, and cells were incubated on ice for an additional 10 min. Precipitated material was collected by centrifugation, and the supernatant was discarded. The tube was washed with 500 µl of 1 M Tris pH 8.0, centrifuged briefly, and the supernatant was discarded. The pellet was vigorously resuspended in 150 µl of 1X Laemlli sample buffer and boiled for 4 min. As a positive control, lysate from HEK293T cells was prepared by sonication in sonication buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl, 0.2 mM

EDTA) and combined with an equal volume of 2X Laemlli sample buffer. Samples (15 μl each) were resolved by SDS-PAGE, transferred to poly(vinylidene difluoride) in CAPS-ethanol pH 10, and probed sequentially with anti-Dicer (Abcam, ab14601) and anti-actin (Abcam, ab8224). Immunoblots were developed with HRP-conjugated anti-mouse antibody and enhanced chemiluminescence (Amersham).

Small-RNA sequencing and analysis

Sequencing libraries. Standard single-turnover reactions were performed with the following modifications. 500 bp body-labeled dsRNA substrates contained 5'-monophosphate termini, and tRNA was omitted from dsRNA dilutions. For each enzyme, separate reactions were performed using dsRNA corresponding to fragments of the genes encoding *Renilla* luciferase and green fluorescent protein. Reactions were performed in a volume of 400 μl, with components at the same concentrations as in the standard reactions. Reactions were incubated at room temperature for 90 s and quenched by addition to an equal volume of phenol supplemented with 1/10 volume 3 M NaCl and 1/74 volume 0.5 M EDTA, pooling reactions performed with the same enzyme but different dsRNA substrates. Total RNA was isolated from dicing reactions by phenol extraction and precipitation. Small-RNA cDNA libraries were prepared as described (Grimson et al., 2008) and sequenced using the Illumina SBS platform.

Analysis. After removing the adaptor sequences, reads representing the small RNAs were collapsed to a non-redundant set, and 14–30 nt sequences were mapped to the appropriate template (i.e., *Renilla* luciferase or GFP), allowing no mismatches and recovering all hits (Table S1). To detect phasing, the frequencies of same-strand distances separating all 23 nt 5'-end pairs mapping to the substrate were calculated separately for each substrate as described (Drinnenberg et al., 2009). The frequencies of opposite-strand distances were calculated similarly, except that the 5' ends of 23 nt reads mapping to the (–)-strand were first converted into the corresponding 5' end on the (+)-strand that was generated by the same RNase III cleavage event (i.e., position *n* on the (–)-strand was mapped to position *n* + 3 on the (+)-strand). Each set of distance frequencies was then normalized to the total number of 5'-end pairs. Same-strand and opposite-strand relative frequencies were averaged between the *gfp* and *Renilla* luciferase substrates. The reported relative frequency of 5'-end pairs is the average of the same- and opposite-strand frequencies.

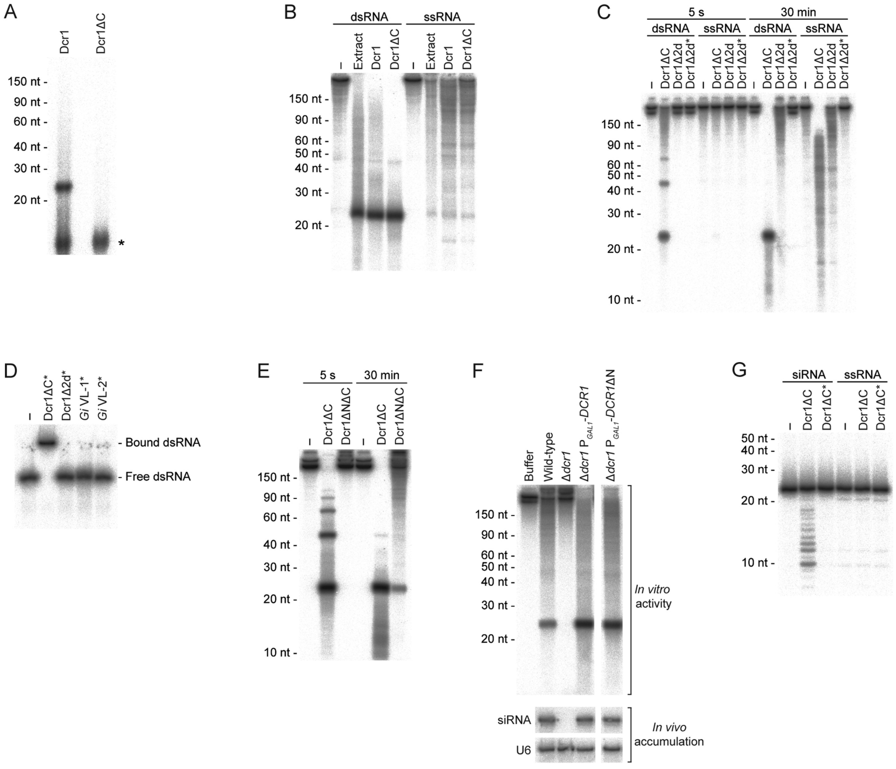


Figure S1. Characterization of Dcr1 variants

(A) Analysis of contaminating RNA in purified *K. polysporus* Dcr1 and Dcr1ΔC. Nucleic acid was isolated from the indicated purified protein by phenol extraction, RNA was 3' end-labeled with cordycepin, and products were resolved by denaturing PAGE. A background band observed in cordycepin-mediated labeling reactions is indicated (*).

(B) Specificity of purified Dcr1 and Dcr1ΔC for dsRNA. Body-labeled 500 bp dsRNA or 500 nt ssRNA was incubated with buffer only (–), extracts prepared from wild-type *K. polysporus* (Extract), or the indicated purified protein under standard single-turnover conditions. Products were resolved by denaturing PAGE.

(C) Activity of *K. polysporus* Dcr1Δ2d, which spans Ser15 to Met260. Body-labeled 500 bp dsRNA or 500 nt ssRNA was incubated for the indicated time with buffer only (–) or the indicated purified protein (Dcr1Δ2d*, Dcr1Δ2d containing the E224Q active-site mutation) under standard single-turnover conditions. Products were resolved by denaturing PAGE.

(D) Binding of Dcr1ΔC and its variants. Trace amounts of body-labeled 500 bp dsRNA was incubated with buffer only (–) or 3 nM of the indicated purified protein (Dcr1ΔC* and Dcr1Δ2d*, variants containing the E224Q active-site mutation; *Gi* VL-1* and *Gi* VL-2*, Dcr1ΔC variants with the E224Q active-site mutation and *Gi*Dicer loop-region substitutions) and separated by native gel electrophoresis.

(E) Activity of *K. polysporus* Dcr1ΔNΔC. Body-labeled 500 bp dsRNA present at 4.67 nM was incubated for the indicated time with buffer only (–) or 1 μM of the indicated purified protein. Products were resolved by denaturing PAGE.

(F) Comparison of *in vitro* activity with *in vivo* product accumulation in strains expressing the corresponding Dcr1 variants. Top: Body-labeled 500 bp dsRNA was incubated with extracts from *S. castellii* strains with the indicated deletions, additions, and replacements. Bottom: An

RNA blot with samples from the same strains was probed for an endogenous siRNA, then reprobed for U6 small nuclear RNA.

(G) Processing of siRNA duplexes by *K. polysporus* Dcr1ΔC. A 23 nt siRNA duplex containing one 5' end-labeled strand (siRNA) or a single-stranded end-labeled 23 nt RNA (ssRNA) was incubated with buffer only (–), wild-type Dcr1ΔC (WT), or an active-site mutant of Dcr1ΔC (Dcr1ΔC*) for 30 min. RNA and protein were present at 1 nM and 30 nM, respectively. Products were resolved by denaturing 20% PAGE.

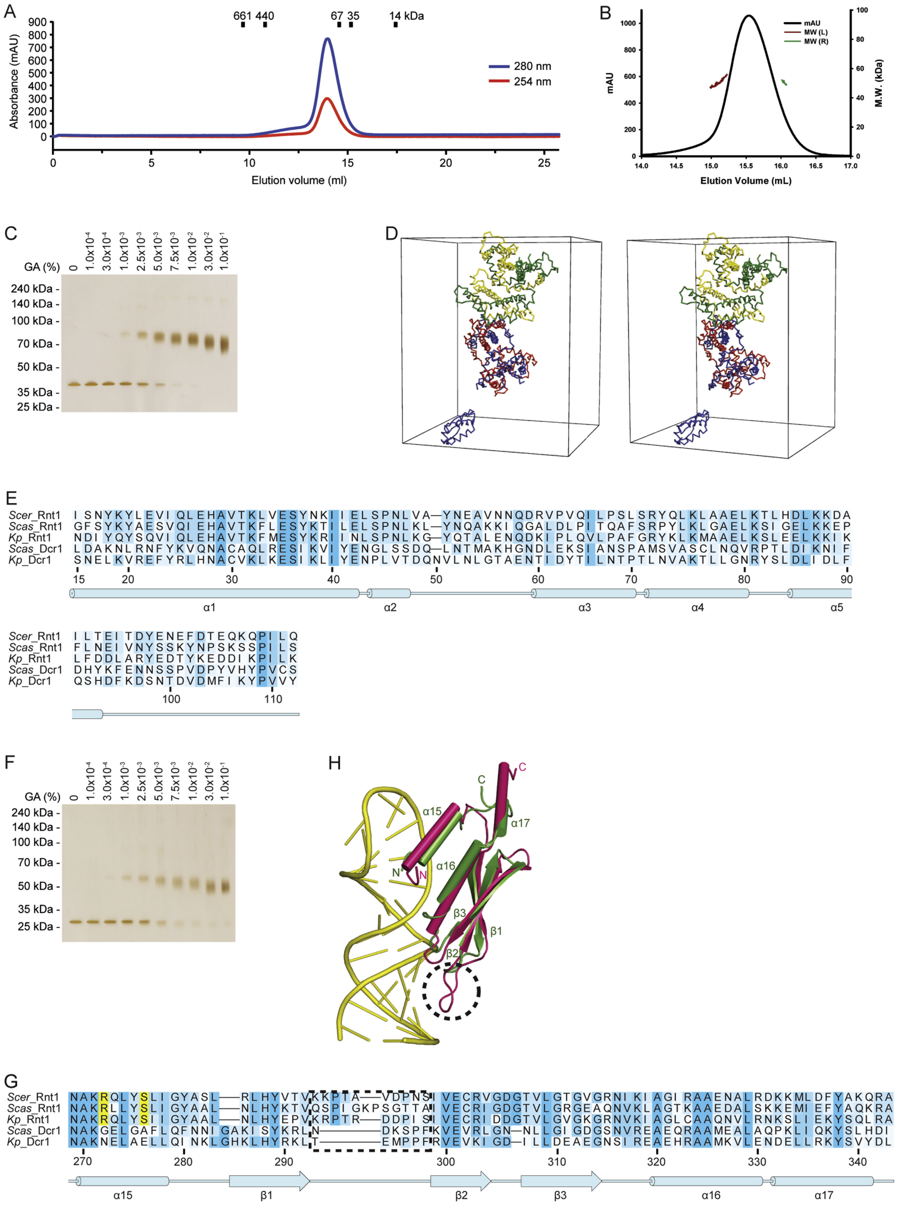


Figure S2. Analysis of Dcr1ΔC sequence and structure

(A) Gel-filtration profile of Dcr1ΔC. Molecular weights of standards are shown above the absorbance trace. The expected molecular weights for the monomeric and dimeric forms of Dcr1ΔC are 39 kDa and 79 kDa, respectively.

(B) Multi-angle light scattering (MALS) analysis of Dcr1Δ2d. The unique fraction from gel filtration (black trace) had a molecular weight of 51 kDa (brown and green traces) over the peak. The expected molecular weights for the monomeric and dimeric forms of Dcr1Δ2d are 28 kDa and 57 kDa, respectively.

(C) Glutaraldehyde crosslinking of Dcr1ΔC. Dcr1ΔC was incubated with the indicated glutaraldehyde (GA) concentration, reactions were analyzed by SDS-PAGE, and proteins were visualized by silver staining. Also shown is the migration of protein standards with the indicated molecular weights. The faster migration of the crosslinked dimer species at high GA concentrations is attributable to the accumulation of intramolecular crosslinks.

(D) Stereo view of the contents of the asymmetric unit in the Dcr1ΔC crystal structure. We could trace four molecules composed of the NTD and RNase III domains (each shown in a different color), as well as one of the four dsRBD1 domains. Two of the remaining dsRBD1 domains were disordered, and the fourth exhibited very weak electron density.

(E) Sequence alignment of N-terminal domains found in RNase III enzymes from representative species in the *Saccharomyces* clade. Annotations of conservation, residue numbers, secondary structure, and species abbreviations are as in Figure 3A.

(F) Glutaraldehyde crosslinking of Dcr1ΔNΔC. Reactions using purified Dcr1ΔNΔC were performed and analyzed as in C. The expected molecular weights for the monomeric and dimeric forms of Dcr1ΔNΔC are 29 kDa and 58 kDa, respectively.

(G) Sequence alignment of Dcr1 dsRBD1 and Rnt1 dsRBD, drawn as in (E). The preference of Rnt1 for AGNN tetraloops is conferred by residues in helix α1 of the dsRBD (Wu et al., 2004), which corresponds to helix α15 in Dcr1 (Figure 2B and Figure S2H). These residues, which are highlighted in yellow, are conserved among RNase III enzymes that recognize AGNN tetraloops but variant in Dcr1 enzymes. The β1-β2 loop that is especially short in the Dcr1 dsRBD1 is boxed.

(H) Superposition of the Dcr1ΔC dsRBD (green) on the Rnt1 dsRBD–hairpin complex (magenta and yellow, respectively; PDB: 1T4L). The β1-β2 loop of Rnt1 dsRBD is circled. The orientation of the Dcr1ΔC dsRBD is as in Figure 2B, central panel. Several residues at the N-terminus of the Dcr1ΔC dsRBD were deleted for clarity.

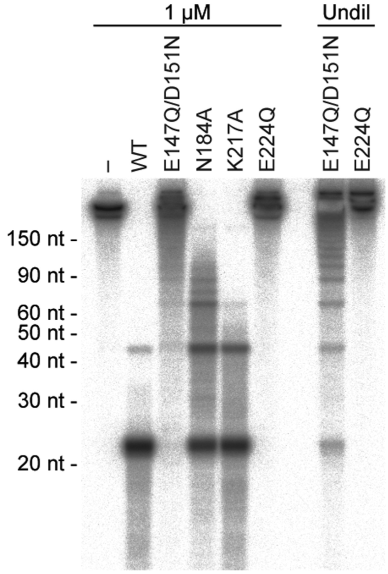


Figure S3. Activity of Dcr1ΔC active-site mutants
Body-labeled 500 bp dsRNA was incubated with buffer only (–), wild-type Dcr1ΔC (WT), or Dcr1ΔC variants with the indicated substitutions. Reactions contained dsRNA present at 140 pM and protein at the indicated final concentration, where undiluted (Undil) E147Q/D151N and E224Q mutants corresponded to 7.5 μM and 29.8 μM, respectively, without BSA. Products were resolved by denaturing PAGE.

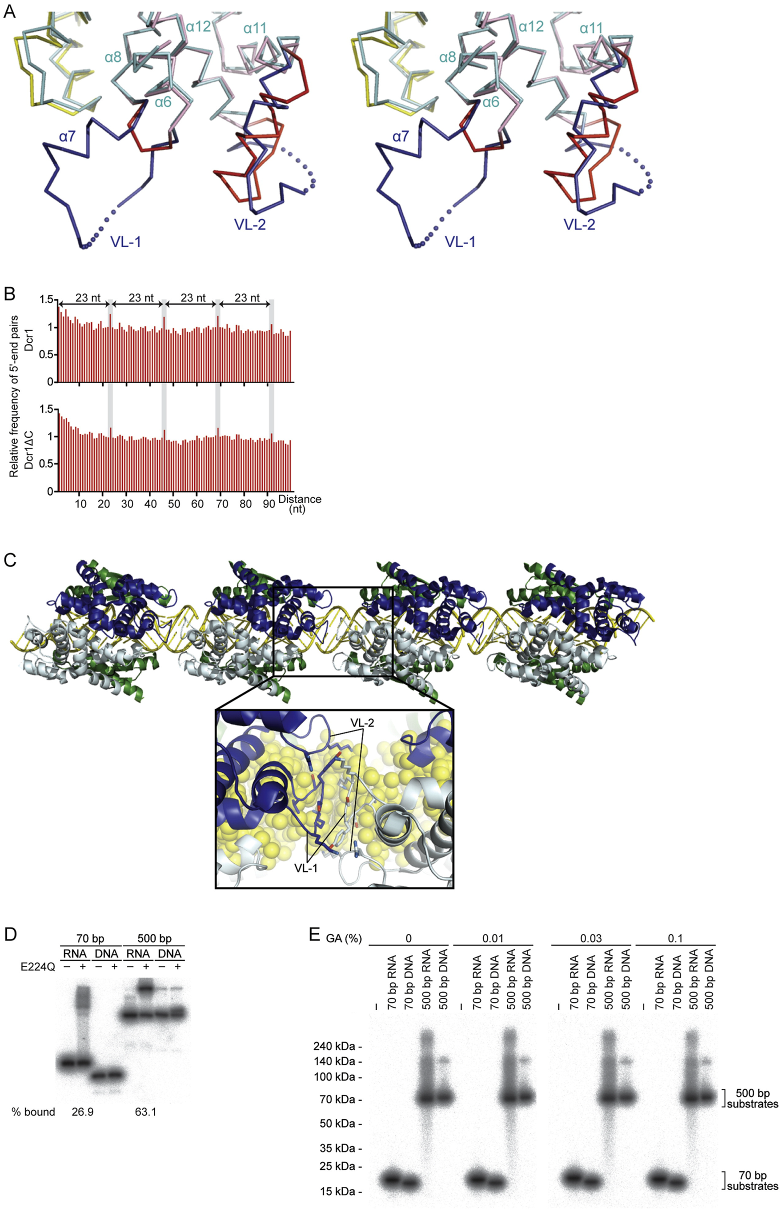


Figure S4. Binding and cleavage within a dsRNA duplex

(A) Design of variable-loop substitutions in Dcr1ΔC. Stereo view of the superposition of the RNase III domain of Dcr1ΔC (cyan) with the RNase IIIa (yellow) and RNase IIIb (pink) domains of *Gi*Dicer. In our loop-swap experiments, either the VL-1 or VL-2 region of the RNase III domain of Dcr1ΔC (dark blue) was replaced with the corresponding region from the RNase IIIb domain of *Gi*Dicer (red). *Gi*Dicer was used for loop-swapping experiments because its structure is known and the RNase IIIb loops are expected to be solvent exposed based on structural modeling with a cognate dsRNA substrate (Macrae et al., 2006). Because the mechanism of *Gi*Dicer does not involve intermolecular interactions between RNase III domains, the RNase IIIb loops are unlikely to contain amino acids that would mediate favorable intermolecular interactions. In contrast, the *Aa*RNase III loops can mediate crystal-packing interactions (C), suggesting that its variable loops might be able to mediate interdimer interactions in solution.

(B) Distribution of the intervals separating the 5' termini of sequenced 23 nt products from *in vitro* dicing reactions using the indicated purified enzymes. Plotted is the relative frequency of each interval, when considering all pairs of reads less than 100 nt apart.
(C) Crystal-packing interactions between *Aa*RNase III dimers. The top view shows the crystal packing between adjacent complexes in the *A. aeolicus* RNase III enzyme (PDB code 2NUG). The bottom view shows a close up of the interface between adjacently positioned RNase III dimers. The color code is the same as in Figure 2A. The 22 nt dsRNA (yellow) forms a pseudo-continuous long dsRNA as a result of packing interactions in the crystal.
(D) Gel-shift analysis of binding reactions used for crosslinking in Figure 4F. BSA (–) or E224Q Dcr1ΔC (+) was incubated with the indicated nucleic acid, and reactions were subjected to native gel electrophoresis.
(E) Analysis of nucleic acid in crosslinking reactions. The RNA and DNA of Figure 4F were radiolabeled to allow their fate to be followed with a phosphorimager. This panel shows that crosslinking did not alter the migration of the RNA on SDS-PAGE and that a majority of the crosslinked bands did not overlap with RNA, even though gel-shift assays of the previous panel (D) indicated that a significant fraction of dsRNA was bound by Dicer. These results confirmed that the crosslinked species observed in Figure 4F represented protein-protein crosslinks.

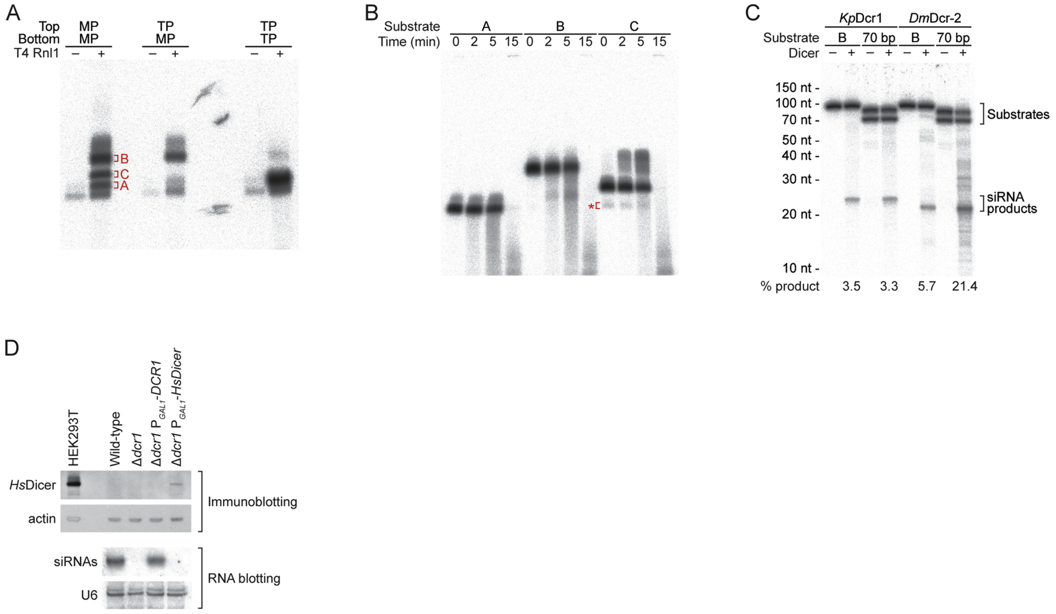
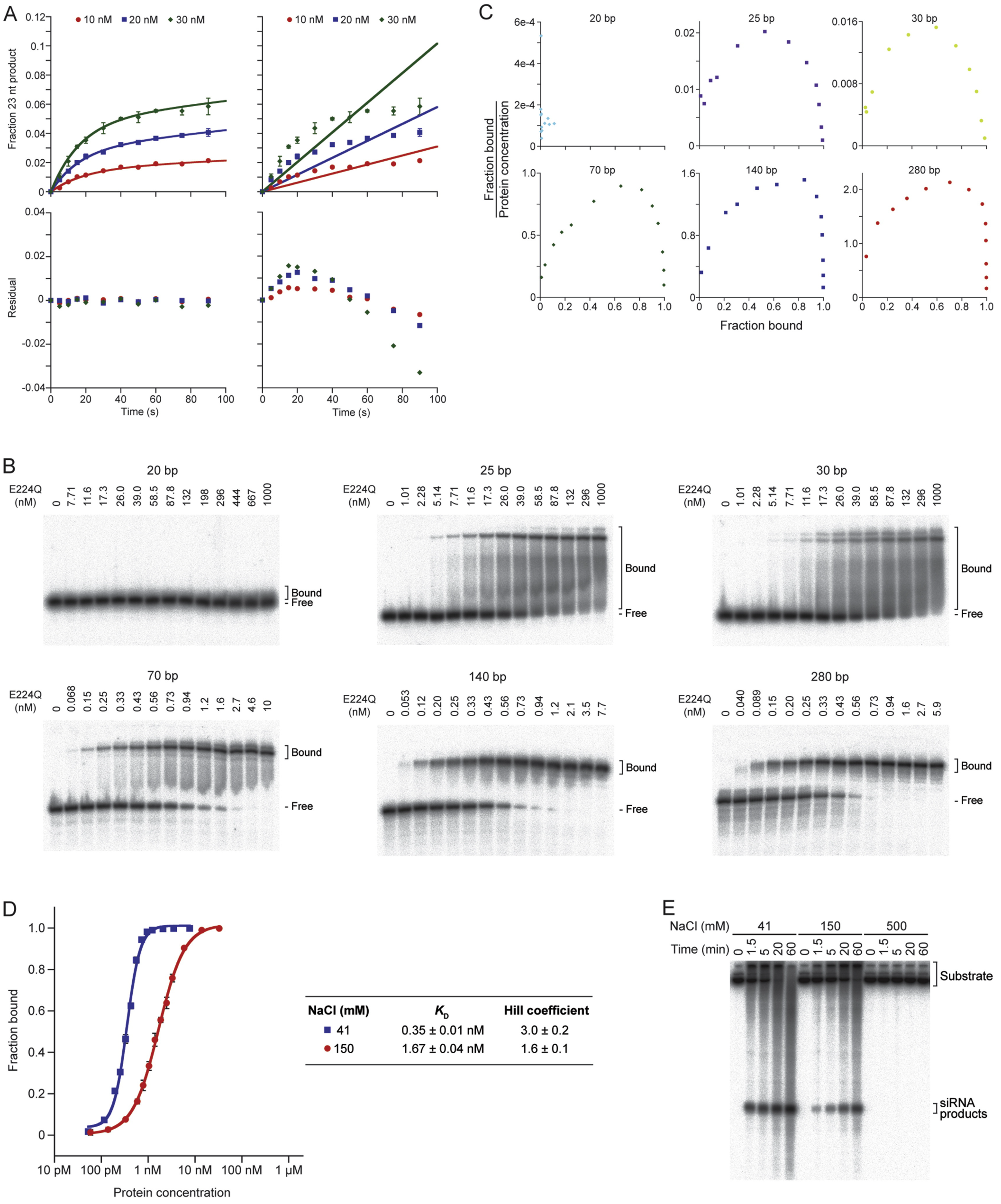


Figure S5. Generation, confirmation, and processing of Dicer substrates
(A) RNA ligation reactions that generated substrates A–C (Figure 5D). Annealed ssRNA duplexes containing the indicated 5'-ends (MP, monophosphate; TP, triphosphate) on the top and bottom strands were incubated without (–) or with (+) T4 RNA Ligase 1 (T4 Rnl1). Products were resolved by denaturing PAGE and distinguished as described in Extended Experimental Procedures.
(B) Limited hydrolysis of substrates A–C. Substrates were diluted in 1X Alkaline Hydrolysis Buffer (Ambion), heated at 90°C for the indicated time, and incubated on ice. Products were analyzed by denaturing PAGE (8% acrylamide, 90% formamide). Contaminating open duplex in the substrate C preparation is indicated (*).
(C) Comparison of Dicer activities on 70 bp substrates. Body-labeled substrates were incubated without (–) or with (+) *K. polysporus* Dcr1ΔC (*Kp*Dcr1) or *D. melanogaster* Dcr-2 (*Dm*Dcr-2) under multiple-turnover conditions (30 nM substrate and 10 nM protein). Substrate B, as in Figure 5D; 70 bp, a perfect duplex without T7-template–encoded ssRNA overhangs. Products were resolved by denaturing PAGE.
(D) Examination of human Dicer activity expressed in *S. castellii*. Top: Immunoblot probing for *H. sapiens* Dicer (*Hs*Dicer) in *S. castellii* strains with the indicated deletions and additions. Lysates prepared from HEK293T cells that endogenously express *Hs*Dicer served as a positive control. Bottom: RNA blot with samples from the same strains was probed for endogenous siRNAs derived from a palindromic RNA (corresponding to substrate E in Figure 5E), then reprobed for U6 small nuclear RNA. Although *Hs*Dicer protein was expressed in *S. castellii*, no siRNAs derived from the palindromic RNA were detected. These results suggest that human Dicer is not able to process the *S. castellii* substrate *in vivo*, which is consistent with the inability of a canonical Dicer to generate siRNAs from this palindromic RNA *in vitro* (Figure 5E). However, additional factors, such as expression level, localization, and interactions with other proteins, might have impaired the activity of human Dicer in this heterologous expression system.



NaCl (mM)	K_D	Hill coefficient
■ 41	0.35 ± 0.01 nM	3.0 ± 0.2
● 150	1.67 ± 0.04 nM	1.6 ± 0.1

Figure S6. Analysis of multiple-turnover kinetics and cooperative binding

(A) Comparison of linear and biphasic fits to the data of Figure 6B. Data were fit to a double-exponential (left) or linear (right) function, and the resulting least-squares fit line (top) and residuals (bottom) were plotted. This analysis confirmed that the data are more appropriately described by a double-exponential function.

(B) Representative gel-shift assays. Trace amounts of body-labeled dsRNA of the indicated length were incubated with increasing amounts of protein and subjected to native gel electrophoresis.

(C) Scatchard plots of the binding data of Figure 6F. The concavity of the plots observed for substrates >20 bp indicates positive cooperativity. In contrast to the full-length enzyme, Dcr1ΔC was less prone to aggregation and could be purified away from contaminating RNA (Figure

S1A), which made it suitable for biochemical studies. Nonetheless, the observed cooperativity of Dcr1ΔC strongly implies that the full-length enzyme also binds cooperatively to dsRNA substrates, because although the deletions that convert full-length Dcr1 to Dcr1ΔC might decrease interdimer interactions and thereby decrease binding cooperativity, they would be unlikely to increase binding cooperativity.

(D) Binding isotherms for E224Q Dcr1ΔC binding to 140 bp dsRNA at low (41 mM) or physiological (150 mM) monovalent ion concentrations. Data points represent average values (n = 3; error bars indicate standard deviation). Solid lines show the best fit to the Hill equation, which produced the dissociation constants (K_D) and Hill coefficients shown to the right.

(E) Product accumulation under multiple-turnover conditions as a function of salt concentration. Body-labeled 500 bp dsRNA present at 14 nM (which corresponded to ~300 nM non-overlapping 23 bp sites) was incubated with 30 nM Dcr1ΔC for the indicated time with the indicated final concentration of NaCl.

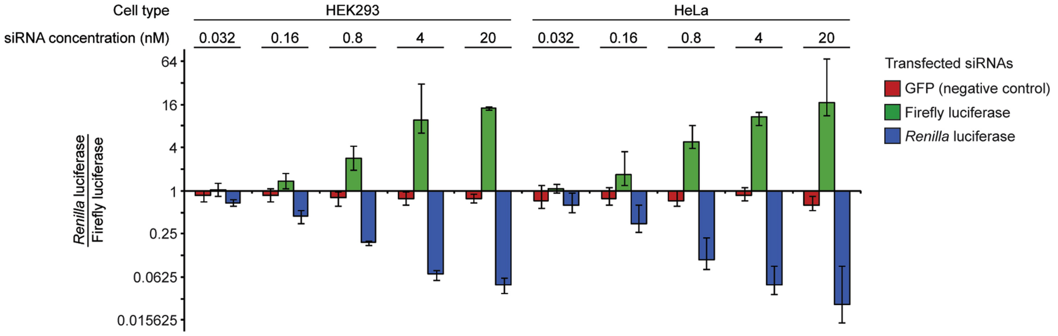


Figure S7. Using siRNAs generated by budding-yeast Dicer for mammalian gene knock-down

Shown is the ratio of *Renilla* luciferase to firefly luciferase following transfection of the indicated cell type with the indicated siRNAs. Plotted are the geometric means, normalized to the geometric means for cells in which no siRNAs were transfected. Error bars represent the largest and the smallest values among 6 replicates (from two independent experiments).

siRNA preparation. Separate reactions were performed using dsRNA corresponding to fragments of the genes encoding *Renilla* luciferase, firefly luciferase, and green fluorescent protein (GFP). Reactions were performed in a volume of 1 ml with 339 nM Dcr1(1-384) and 10 ng/µl dsRNA. Reactions were incubated at room temperature for 45 min and quenched by addition of 1/5 volume 50 mM EDTA supplemented with 1.5 M NaCl. After addition of 1 ml dsRNA Storage Buffer, total RNA was isolated by phenol extraction and precipitation, and siRNAs were enriched for using RNA Purification Column 2 (Genlantis). To confirm the production of siRNA duplexes, a portion of the preparation was analyzed by native 15% PAGE.

Cell culture and luciferase assays. One day prior to transfection, HeLa or HEK293 cells were seeded into 24-well plates at a density of 0.5–1 x 10^5 cells/well in 500 µl of DMEM supplemented with 10% FBS. Cells were transfected in triplicate with 100 ng each of pIS0 and pIS1 (expressing firefly and *Renilla* luciferase genes, respectively) and 5-fold serial dilutions (20 nM to 32 pM) of siRNA products. Twenty-two hours later, cells were washed with PBS, and *Renilla* and firefly luciferase levels were determined using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's directions.

Table S1. Analysis of small-RNA libraries

Numbers in parentheses indicate the percent of reads compared to total number of substrate-matching reads.

	K. polysporus Dcr1		*K. polysporus* Dcr1ΔC	
All substrate-matching reads	290550	(100)	839110	(100)
23 nt substrate-matching reads	232873	(80.1)	588945	(70.2)
GFP				
All reads	102732	(35.4)	314579	(37.5)
23 nt reads	74065	(25.5)	202214	(24.1)
Luciferase				
All reads	187818	(64.6)	524531	(62.5)
23 nt reads	158808	(54.7)	386731	(46.1)

Table S2. Crystallographic data statistics for *K. polysporus* Dcr1ΔC and Dcr1Δ2d

Structure	Dcr1ΔC		Dcr1Δ2d
Crystal	Native	SeMet	Native
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (a, b, c [Å])	101.0, 113.0, 135.7	101.0, 113.0, 134.7	58.9, 96.4, 101.4
Wavelength (Å)	0.9795	0.9792	0.9792
Resolution (Å)[a]	50.00–2.30 (2.38–2.30)	50.00–3.50 (3.56–3.50)	50.00–1.97 (2.04–1.97)
R_{sym} [a]	0.095 (0.636)	0.157 (0.540)	0.094 (0.458)
$I/\sigma(I)$ [a]	27.3 (2.8)	25.5 (9.8)	16.7 (2.1)
Completeness (%)[a]	99.7 (97.7)	100.0 (100.0)	97.9 (96.1)
Unique reflections[a]	69.722 (6718)	38,291 (1758)	40,348 (406)
Redundancy[a]	10.0 (8.8)	7.7 (7.8)	5.6 (4.1)
Structural Refinement			
Resolution (Å)	43.41–2.29		48.21–1.97
Number of reflections	64,882		40,311
$R_{\mathrm{work}}/R_{\mathrm{free}}$	17.51/21.91 (20.72/27.63)		19.30/23.75 (34.95/39.57)
Number of Atoms			
Protein	8157		3699
Ion	4		-
Water	467		262
Average B-factors (Å^2)			
Protein	46.21		26.04
Ion	48.03		-
Water	41.18		30.47
Rmsd Values			
Bond lengths (Å)	0.010		0.007
Bond angles (º)	0.974		0.947

[a] Values for the highest resolution shell are in parentheses.

Table S3. Yeast strains used and generated in this study

Strain	Genotype	Species	Reference
KpolWT	Wild-type	*K. polysporus* DSM70294	(Scannell et al., 2007)
DPB277	*MATα hoΔ ura3*::*EGFP(S65T)-KanMX6 Flag*$_3$*-AGO1*	*S. castellii* CBS4310	This study
DPB278	*MATα hoΔ ura3*::*EGFP(S65T)-KanMX6 Flag*$_3$*-AGO1 dcr1::ScerURA3*	*S. castellii* CBS4310	This study
DPB406	*MATα hoΔ ura3*::*EGFP(S65T)-KanMX6 Flag*$_3$*-AGO1 dcr1::DCR1*	*S. castellii* CBS4310	This study
DPB437	*MATα hoΔ ura3*::*EGFP(S65T)-KanMX6 Flag*$_3$*-AGO1 dcr1::DCR1ΔC*	*S. castellii* CBS4310	This study
DPB005	*MATα hoΔ ura3-1*	*S. castellii* CBS4310	(Drinnenberg et al., 2009)
DPB318	*MATα hoΔ ura3-1 dcr1Δ*	*S. castellii* CBS4310	(Drinnenberg et al., 2009)

Table S4. Plasmids generated in this study

Plasmid	**Description**
pRSF-KpDcr1	*E. coli* expression plasmid, *K. polysporus DCR1*
pRSF-KpDcr1ΔC	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355)
pRSF-KpDcr1ΔC(E147Q/D151N)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with E147Q/D151N
pRSF-KpDcr1ΔC(N184A)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with N184A
pRSF-KpDcr1ΔC(K217A)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with K217A
pRSF-KpDcr1ΔC(E224Q)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with E224Q
pRSF-KpDcr1Δ2d	*E. coli* expression plasmid, *K. polysporus DCR1*(15–260)
pRSF-KpDcr1Δ2d(E224Q)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–260) with E224Q
pRSF-KpDcr1ΔNΔC	*E. coli* expression plasmid, *K. polysporus DCR1*(107–355)
pRSF-KpDcr1(1–384)	*E. coli* expression plasmid, *K. polysporus DCR1*(1–384)
pRSF-KpDcr1ΔC(*Gi* VL-1)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with H123–N143 replaced by P642–Y645 of *Gi*Dicer
pRSF-KpDcr1ΔC(E224Q, *Gi* VL-1)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with E224Q and with H123–N143 replaced by P642–Y645 of *Gi*Dicer
pRSF-KpDcr1ΔC(*Gi* VL-2)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with N195–M215 replaced by P697–D714 of *Gi*Dicer
pRSF-KpDcr1ΔC(E224Q, *Gi* VL-2)	*E. coli* expression plasmid, *K. polysporus DCR1*(15–355) with E224Q and with N195–M215 replaced by P697–D714 of *Gi*Dicer
pYES2.1-ScDcr1	2-micron plasmid, *S. castellii DCR1* under *GAL1* promoter
pYES2.1-ScDcr1ΔC	2-micron plasmid, *S. castellii DCR1*(17–356) under *GAL1* promoter
pYES2.1-ScDcr1ΔN	2-micron plasmid, *S. castellii DCR1*(111–610) under *GAL1* promoter
pAG416Gal-Dicer	*CEN* plasmid, *H. sapiens DICER* under *GAL1* promoter (Suk et al., 2011)

Supplemental References

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Collaborative Computational Project (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In Methods in Enzymology, C.W. Carter, and R.M. Sweet, eds. (San Diego, Academic Press), pp. 472-494.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A *47 (Pt 2)*, 110-119.

Liu, L., Botos, I., Wang, Y., Leonard, J.N., Shiloach, J., Segal, D.M., and Davies, D.R. (2008). Structural basis of toll-like receptor 3 signaling with double-stranded RNA. Science *320*, 379-381.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, C.W. Carter, and R.M. Sweet, eds. (San Diego, Academic Press), pp. 307-326.

Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. Nucleic Acids Res *35*, e60.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A *104*, 8397-8402.

Suk, K., Choi, J., Suzuki, Y., Ozturk, S.B., Mellor, J.C., Wong, K.H., Mackay, J.L., Gregory, R.I., and Roth, F.P. (2011). Reconstitution of human RNA interference in budding yeast. Nucleic Acids Res.

Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr *56*, 1622-1624.

Weeks, C.M., and Miller, R. (1999). The design and implementation of *SnB* version 2.0. Journal of Applied Crystallography *32*, 120-124.

Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004). Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. Proc Natl Acad Sci U S A *101*, 8307-8312.

Chapter 4

Structure of Yeast Argonaute with Guide RNA

Kotaro Nakanishi[1,]*, David E. Weinberg[2,3,]*, David P. Bartel[2,3], Dinshaw J. Patel[1]

[1] Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065
[2] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142
[3] Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

* These authors contributed equally to this work.

D.E.W. performed the biochemical experiments under the supervision of D.P.B. K.N. performed the structural experiments under the supervision of D.J.P. All authors designed the study and wrote the manuscript.

Published as:
Nakanishi K*, Weinberg DE*, Bartel DP, Patel DJ (2012). Structure of yeast Argonaute with guide RNA. *Nature* 486:368–74.
** indicates equal contributions*

The RNA-induced silencing complex, comprising Argonaute and guide RNA, mediates RNA interference. Here we report the 3.2 Å crystal structure of *Kluyveromyces* Argonaute (*Kp*AGO) fortuitously complexed with guide RNA originating from small-RNA duplexes autonomously loaded by recombinant *Kp*AGO. Despite their diverse sequences, guide-RNA nucleotides 1–8 are positioned similarly, with sequence-independent contacts to bases, phosphates and 2'-hydroxyl groups pre-organizing the backbone of nucleotides 2–8 in a near–A-form conformation. Compared with prokaryotic Argonautes, *Kp*AGO has numerous surface-exposed insertion segments, with a cluster of conserved insertions repositioning the N domain to enable full propagation of guide–target pairing. Compared with Argonautes in inactive conformations, *Kp*AGO has a hydrogen-bond network that stabilizes an expanded and repositioned loop, which inserts an invariant glutamate into the catalytic pocket. Mutation results and analogies to Ribonuclease H indicate that insertion of this glutamate finger completes a universally conserved catalytic tetrad, thereby activating Argonaute for cleavage.

RNA interference (RNAi) is a eukaryote-specific gene-silencing pathway triggered by double-stranded RNA (dsRNA) (Bartel, 2004; Meister and Tuschl, 2004; Malone and Hannon, 2009). In this pathway, the RNase III enzyme Dicer first cleaves the dsRNA trigger into small interfering RNAs (siRNAs), which have 5'-monophosphates and pair to each other with 2-nucleotide 3' overhangs (Bernstein et al., 2001; Elbashir et al., 2001a; Hutvagner et al., 2001). The siRNA duplex is then incorporated into the effector protein Argonaute (AGO), whereupon one of the strands (designated the passenger strand) is cleaved (Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). After the cleaved passenger strand is discarded, the resulting ribonucleoprotein complex (designated the RNA-induced silencing complex, or RISC) uses the remaining siRNA strand (designated the guide strand) to specify interactions with target RNAs (Tomari and Zamore, 2005; Carthew and Sontheimer, 2009). If sequence complementarity between guide and target is extensive, AGO again catalyzes cleavage, resulting in ‘slicing’ of the target RNA (Song et al., 2004).

The first structures of full-length AGOs were of prokaryotic proteins from *Pyrococcus furiosus* (*Pf*AGO) (Song et al., 2004) and *Aquifex aeolicus* (*Aa*AGO) (Yuan et al., 2005). Early structures revealed that the PIWI domain adopts an RNase H–like fold, thereby implicating AGO as the ‘slicer’ enzyme that mediates RNAi (Liu et al., 2004; Parker et al., 2004; Song et al., 2004). Because these prokaryotic enzymes bind 5'-phosphorylated guide DNAs in preference to their RNA counterparts (Ma et al., 2005; Yuan et al., 2005), subsequent structures featured the binary complex of *Thermus thermophilus* Ago (*Tt*AGO) with guide DNA (Wang et al., 2008b) and ternary complexes with target RNAs of varying length (Wang et al., 2008a; Wang et al., 2009). These studies shed light on the nucleation, propagation and cleavage steps of the AGO catalytic cycle (Wang et al., 2009; Parker, 2010). However, the physiological role of prokaryotic AGOs is enigmatic; the origin of the guide DNA is unknown and the host bacteria lack recognizable components of the RNAi pathway (Makarova et al., 2009). Therefore, attention has turned to eukaryotic AGOs, which utilize RNA guides and have protein-binding partners not found in bacteria (Meister et al., 2005). Eukaryotic AGOs are also larger than prokaryotic AGOs, primarily because of additional insertion elements of unknown structure and function. Previous studies determined the structures of individual domains and the MID-PIWI lobe within eukaryotic AGO (Lingel et al., 2003; Yan et al., 2003; Ma et al., 2004; Boland et al., 2010; Frank et al., 2010; Boland et al., 2011), but structural characterization of the entire protein has remained a challenge.

Although *Saccharomyces cerevisiae* lacks RNAi, some closely related budding-yeast species were recently shown to have retained RNAi, thereby offering fresh possibilities for the study of the eukaryotic pathway (Drinnenberg et al., 2009). We previously determined the structure and mechanism of Dicer from the budding yeast *Kluyveromyces polysporus* (Weinberg et al., 2011) and thus turned our attention to the AGO of this species.

Cleavage activity of budding-yeast AGO

K. polysporus AGO (Ago1) has the four conserved domains (N, PAZ, MID, PIWI) and two linker regions (L1, L2) found in other AGOs (Figure 1a). It also has an N-terminal extension, predicted to be disordered, which we removed to facilitate crystallization. The resulting protein, referred to as *Kp*AGO, can substitute for the full-length protein when reconstituting RNAi in *S. cerevisiae* (Figure 1b).

*Kp*AGO and other budding-yeast AGOs have acidic side chains at the three positions corresponding to active-site residues in slicing-competent AGOs (Rivas et al., 2005) (Supplementary Figure 1), which suggested that *Kp*AGO might also cleave target RNAs. Indeed, after incubation with a single-stranded guide RNA, recombinant *Kp*AGO cleaved a matched target RNA at the expected position (Figure 1c). To examine whether slicing occurs *in vivo*, we performed degradome sequencing from another RNAi-containing yeast, *Saccharomyces castellii*. Degradome sequencing identifies polyadenylated RNAs containing 5'-monophosphates, including products of AGO-catalyzed slicing (Addo-Quaye et al., 2008; German et al., 2008). Many *AGO1*-dependent degradome tags mapped to Y'-element transcripts (major targets of *S.*

castellii RNAi (Drinnenberg et al., 2009)) and tended to pair to endogenous siRNAs in the register implicating cleavage across from positions 10–11 of the guide RNA, which was diagnostic of slicing (Elbashir et al., 2001b) (Supplementary Figure 2). These results indicating that budding-yeast AGO functions as a slicer during endogenous RNAi add to the *in vitro* results, establishing *Kp*AGO as a eukaryotic slicer suitable for structure-function analyses.

Structural architecture of eukaryotic AGO

We crystallized *Kp*AGO purified from *E. coli*. Extensive screening eventually identified several crystals that were free of twinning, one of which diffracted to 3.2 Å resolution. A crystal of selenomethionine-substituted *Kp*AGO yielded reflections suitable for phasing by single-wavelength anomalous dispersion (Supplementary Table 1; representative electron density is shown in Supplementary Figure 3).

The overall structure of *Kp*AGO resembles the bilobal architecture of its prokaryotic counterparts but with expansions throughout the protein (Figure 2a and Supplementary Figure 4). Of the 19 insertion segments not found in prokaryotic AGOs (Song et al., 2004; Yuan et al., 2005; Wang et al., 2008b), 11 were conserved segments (cS) found in all eukaryotic AGOs, albeit with some differences in secondary structure and/or length, whereas the remaining eight were variable segments (vS) found in only some eukaryotic AGOs (Supplementary Figure 1). All insertion segments are on the exterior, thereby generating new surfaces for interactions with AGO-binding proteins.

Autonomously loaded guide RNA

After modeling the *Kp*AGO protein, the *F*o–*F*c map revealed continuous residual electron density lying along the nucleic acid–binding channel (Figure 2b). This unanticipated density resembled that of an oligonucleotide and could be fit well with an RNA octamer (Figure 2c and Supplementary Figure 5a). Analysis of end-labeled polynucleotides extracted from soluble and crystalline *Kp*AGO confirmed the presence of small RNAs (Figure 2d and Supplementary Figure 5b), the high-throughput sequencing of which identified a diverse population with a bimodal length distribution centering at 12 and 17 nucleotides (Supplementary Table 2 and Supplementary Figures 5c and 5d).

The location of the small RNAs within the nucleic acid–binding channel suggested that they might represent functional guide RNAs. Supporting this interpretation, co-purifying RNAs had two features of budding-yeast guide RNAs: enrichment for 5' uridine (Figure 2e) and presence of 5' monophosphate, indicated by both electron density (Figure 2c) and a phosphatase-sensitive block of 5'-end labeling (Supplementary Figure 5e) (Drinnenberg et al., 2009). Confirming guide-RNA function, our *Kp*AGO preparation sliced an RNA containing a site complementary to a co-purifying 17-nucleotide RNA comprising ~0.1% of our sequencing reads (Figure 2f and Supplementary Figure 5c). Slicing was at the anticipated linkage and sensitive to mismatches to guide nucleotides 10–11. Reactions displayed initial burst kinetics, as observed previously for metazoan AGOs (Haley and Zamore, 2004; Martinez and Tuschl, 2004; Rivas et al., 2005; Ameres et al., 2007; Forstemann et al., 2007), although addition of Triton enabled sustained product formation (Supplementary Figure 5f), perhaps by facilitating a conformational change that promotes product release.

Most co-purifying RNAs mapped to the *Kp*AGO expression plasmid (Figure 2e and Supplementary Figure 6) in a manner suggesting origins from siRNA-like duplexes loaded into *Kp*AGO with passenger-strand cleavage (Supplementary Figure 7). For such loading to occur, *Kp*AGO must be able to load siRNA duplexes in the absence of RISC-loading factors. Indeed, purified *Kp*AGO incubated with an siRNA duplex generated products diagnostic of passenger-strand cleavage (Figure 2g) and formed active RISC able to slice a cognate target RNA (Figure 2h). Loading was more efficient with duplex than with single-stranded guide and occurred asymmetrically in a manner consistent with preference for 5' uridine on the guide strand (Figure 2h and Supplementary Figures 8a–c).

We conclude that *Kp*AGO can autonomously load an siRNA duplex, lose the passenger strand and then slice targets. This conclusion counters the prevailing view that loading of siRNA duplexes to form functional RISC requires RISC-loading factors (Carthew and Sontheimer, 2009). We suspect that some other AGOs can also autonomously load siRNA duplexes and that reports to the contrary resulted from assaying target-RNA slicing under conditions in which AGO retained inhibitory passenger-strand fragments. Autonomous loading explains how *Kp*AGO RISC fortuitously formed in the absence of other RNAi proteins. In contrast to previous preparations of AGO complexes used for structural studies (Wang et al., 2008b), the formation of *Kp*AGO RISC through loading of a duplex resembles the physiological RISC-assembly pathway. From this perspective, the *Kp*AGO structure reflects the natural state of eukaryotic RISC.

Organization of the guide RNA

Electron density corresponding to the base of nucleotide 1 was smaller than that corresponding to most other positions (Supplementary Figure 9), which agreed with our sequencing results showing that *Kp*AGO-bound RNAs were diverse but enriched for a 5' uridine (Figure 2e). Therefore, we modeled the first nucleotide as uridine and the next seven as adenine (the generic nucleotide used to minimize bias during refinement (Frazao et al., 2006)) and refined the final structure as the *Kp*AGO–pUAAAAAAAp binary complex (Supplementary Table 1 and Supplementary Figure 10).

The guide-strand nucleotides 1–8 run along the nucleic acid–binding channel, from the MID domain to the L2 domain. These nucleotides, including their bases, have electron-density quality resembling that of the *Kp*AGO protein, even though this density represents a composite of thousands of different RNAs. Thus, for this segment of the guide RNA, known as the seed region, diverse RNA sequences are all presented in essentially the same orientation. The electron-density disappeared after the ninth nucleotide (Figures 2b and 2c), even though most co-purifying RNAs were longer than 9 nucleotides (Supplementary Figure 5d). This density loss suggests that guide-RNA 3' halves are either disordered or adopt diverse sequence-specific conformations. In addition, the PAZ domain is not well ordered, as observed in *Tt*AGO complexes in which the PAZ domain has released the 3' end of the guide (Wang et al., 2009), consistent with the idea that *Kp*AGO holds the guide RNA without assistance from the PAZ domain.

Like prokaryotic AGOs (Parker et al., 2004, 2005; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009), *Kp*AGO recognizes the 5' phosphate of the guide, the notable difference being that *Kp*AGO uses the ammonium group of Lys939 rather than a divalent cation (despite Mn^{2+} in the crystallization buffer) to neutralize the negative charge resulting from the close juxtaposition of the C-terminal carboxylate and phosphates 1 and 3 (Figures 3a and 3b). The inserted C-terminus is anchored by Lys939 and Lys943 (Figure 3a), with mutation of either residue impeding guide-RNA binding in *Drosophila* AGO1 (Boland et al., 2011). Another distinct facet involves Arg1183, which hydrogen bonds with the C-terminal carboxylate and phosphate 4 (Figure 3c). In the free *Nc*QDE-2 MID-PIWI structure (Boland et al., 2011) Arg1183 is in a disordered loop, suggesting that guide RNA recruits Arg1183 to the 5'-phosphate–binding pocket. Notably, conservation of Lys939 and Arg1183 is restricted to eukaryotic AGOs (Supplementary Figure 1).

The Asn897 main-chain amide interacts with the O2 carbonyl of the uridine at position 1 (Figure 3a). Because analogous interactions with O2 of cytidine and N3 of purines would be isosteric, this hydrogen bond cannot explain the preference for a 5' uridine. The preference might instead be attributed to the relatively weak stacking energy of uridine, which would facilitate the requisite flipping out of nucleotide 1 during siRNA loading.

*Kp*AGO interacts with phosphates of the seed region primarily using contacts homologous to those observed in prokaryotic AGO complexes (Ma et al., 2005; Parker et al., 2005; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009) (Figure 3c). Structures of prokaryotic complexes, however, have not revealed intermolecular contacts to the guide-RNA 2'-OH groups. We find that *Kp*AGO forms hydrogen bonds with most 2'-OH groups of the seed, using main-chain atoms at positions 2, 5 and 6 and hydroxyl groups of Thr1186 and Tyr681 at positions 4 and 7, respectively (Figure 3d). We also observe an intra-RNA hydrogen bond between the 2'-OH group at position 3 and O4' at position 4, a type of interaction proposed to facilitate base-pair fluctuations in A-form RNA helices (Pan and MacKerell, 2003). A second intra-RNA hydrogen bond involves the 2'-OH group at position 1 and a non-bridging oxygen of phosphate 2, as previously observed in *Af*PIWI–siRNA complex structures (Ma et al., 2005; Parker et al., 2005).

To examine the contributions of guide-strand 5' phosphate and 2'-OH groups, we monitored autonomous loading and passenger-strand cleavage of modified siRNA duplexes. Removing the monophosphate or substituting all guide-strand 2'-OH groups with 2'-H (deoxy) greatly impaired activity (Figure 3e and Supplementary Figure 11), consistent with observations in transfected human cells (Schwarz et al., 2002; Chiu and Rana, 2003). To learn more about the 2'-OH groups contributing to this effect, we compared guide RNAs with deoxy substitutions at positions 1, 2–8, 9–14, 15–21 and 22–23. Substitution of the 2'-OH group at position 1 enhanced activity (perhaps by facilitating flipping out of nucleotide 1), whereas substitutions in all other regions impaired activity (Figure 3e and Supplementary Figure 11). Deoxy substitution at positions 2–8 impaired activity to a similar degree as at positions 22–23, which are presumably recognized by the PAZ domain (Lingel et al., 2003; Yan et al., 2003; Ma et al., 2004). Thus, the 2'-OH groups within the seed region contribute to duplex loading or passenger-strand cleavage. Nonetheless, greater effects were observed at positions 9–14 and 15–21, the understanding of which will require structural studies of additional states along the eukaryotic RISC-assembly pathway.

Together, contacts to the phosphate and 2'-OH groups maintain the sugar–phosphate backbone of the single-stranded guide-RNA seed in a near–A-form conformation resembling that of the siRNA duplex (Figure 3f). Maintaining this conformation pre-organizes the seed backbone for pairing to the target, as anticipated from studies of microRNA targeting (Bartel, 2004) and supported by structural and biophysical studies (Ma et al., 2005; Parker et al., 2005; Parker et al., 2009). Also as anticipated, the bases of the seed nucleotides are stacked, with Watson–Crick faces (particularly those of nucleotides 2–4) displayed to solvent and accessible to nucleate pairing to target RNA (Figure 3g).

The surprising feature of the guide-RNA conformation was the tilting of the bases away from the orientation required for helical pairing (Figure 3f). *Kp*AGO makes hydrophobic contacts with the bases at positions 2, 5 and 6 while anchoring the sugar–phosphate backbone (Figures 3c and 3d). Base 2 stacks on Tyr932 (Figure 3d), which is conserved as Tyr or Thr in eukaryotic AGOs (Supplementary Figure 1) and thus might represent a conserved hydrophobic interaction that facilitates the flipping out of nucleotide 1 by preventing its stacking on base 2. As observed in structures of prokaryotic AGO complexes (Ma et al., 2005; Parker et al., 2005; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009), base 2 is recognized at N3 (purines) or O2 (pyrimidines) by the side chain of Asn935 (Figure 3d), which is conserved throughout all AGOs. Bases 5 and 6 are surrounded by a hydrophobic pocket comprising Ile682, Ala686, Leu1147 and Lys1148 (Figure 3d). Bases 3 and 4 make no contact with *Kp*AGO but are nonetheless tilted because of continuous stacking of the seed bases (Figures 3c and 3d). Untilting of the seed stack, which would accompany nucleation of target pairing at positions 2–4, might disfavor contacts to Ile682 and neighboring residues, thereby facilitating repositioning of $\alpha16$, a helix that would otherwise block full seed pairing. Such changes in base tilting and $\alpha16$ might communicate the presence of target RNA.

Potentially unobstructed guide–target pairing

To compare the architectures of eukaryotic and prokaryotic AGOs, we structurally aligned each domain of *Kp*AGO on its *Tt*AGO counterpart. Except for the N domain, each of the domains superimposed well (Supplementary Figure 12). The structural difference between the N domains is attributed to cS1, cS3 and vS2 (Figures 4a and 4b). cS1 and cS3 cluster together with cS7 and cS10 such that they bury a space observed in prokaryotic AGO structures and concomitantly lengthen the nucleic acid–binding channel (Song et al., 2004; Yuan et al., 2005; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009) (Figures 4c and 4d). These insertion segments interact with the L2 and PIWI domains through a hydrogen-bond network involving residues that are conserved throughout eukaryotic AGOs (Supplementary Figures 1 and 13a), which suggests that an extended nucleic acid–binding channel is a feature common to eukaryotic AGOs.

In all crystallized conformations of *Tt*AGO, the N domain blocks the channel and prevents propagation of guide–target pairing beyond position 16 (Wang et al., 2009) (Figure 4c and Supplementary Figures 13c and 13d). In addition to lengthening the nucleic acid–binding channel, the cS1/3/10 cluster positions the *Kp*AGO N domain such that a slight widening of the channel would allow pairing to propagate to the 3' end of the guide RNA (Figure 4d and Supplementary Figure 13b). The potential for unobstructed propagation of guide–target pairing is consistent with the prevalence of pairing throughout the 3' region of plant small RNAs that guide target cleavage (Mallory et al., 2004) and the contribution of such pairing to the stability of guide–target association *in vitro* (Ameres et al., 2007).

Glu1013 completes a catalytic tetrad

When comparing the structures of *Kp*AGO and the free *Nc*QDE-2 MID-PIWI lobe (Boland et al., 2011), we observed striking differences in loops L1 and L2 (Supplementary Table 3). In *Kp*AGO, loop L2 expands by partial unfolding of α25 (Figure 5a) and packs into a cavity, such that the invariant Glu1013 side chain inserts into the catalytic pocket, near the three Asp residues of the active site (Figure 5b). This conformation is enabled by the movement of loop L1, which otherwise blocks access to the catalytic pocket (Figure 5c). Opening of the loop L1 gate in *Kp*AGO is accompanied by a conformational transition of cS11 and hydrophobic packing between aliphatic side chains on loop L1 and cS11. Notably, deletion of cS11 from *Drosophila* AGO1 inhibits guide RNA-binding and abolishes silencing activity (Boland et al., 2011).

The plugged-in conformation, in which the Glu1013 finger is inserted into the catalytic pocket, is stabilized by an extensive hydrogen-bond network, with Glu1013 bridging His977 and Arg1045, and loop L2 main-chain atoms interacting with His977, Arg1045 and Glu1060 (Figure 5b). These four residues are conserved throughout eukaryotic AGOs (and even most of the PIWI clade; Supplementary Figure 1). Glu1013, Arg1045 and Glu1060 are also conserved throughout prokaryotic AGOs, prompting a search for a similar plugged-in conformation in the available structures (Song et al., 2004; Yuan et al., 2005; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2009; Boland et al., 2011). We found both plugged-in and unplugged conformations of *Tt*AGO, with striking parallels to the eukaryotic hydrogen-bond network and correlated loop movements (Figures 5a–f). The plugged-in conformation was observed only in complexes in which the PAZ domain released the 3' end of the guide and *Tt*AGO assumed its catalytically active state (Supplementary Table 3). In contrast, the inactive states—either those of the apo proteins or complexes in which the PAZ domain engages the guide 3' end—resembled the unplugged conformation. These observations suggest that the plugged-in conformation of loop L2 is correlated with release of the 3' end of the guide and formation of active RISC. Confirming the functional importance of the plugged-in conformation, mutation of any of the four residues of the *Kp*AGO hydrogen-bond network impaired RNAi (Figure 5g).

The structures of ternary *Tt*AGO complexes in the plugged-in conformation show the position of loop L2 in the context of guide and target strands. *Tt*AGO loop L2 interacts with the guide DNA at positions 11–15 (Supplementary Table 3) (Wang et al., 2009). Moreover, the carboxyl group of the glutamate finger approaches both the 2'-OH of the nucleotide adjacent to the scissile phosphate and one of the two active-site divalent metal ions (Figure 5h), which suggests that the glutamate finger might act as a catalytic residue. Indeed, simultaneous coordination of the analogous 2'-OH and metal ion is the role of Glu109 in the 'DEDD' catalytic tetrad at the active site of *Bacillus halodurans* RNase H1 (Nowotny et al., 2005). Although the PIWI domain of AGO has an RNase H fold, only a conserved 'DDX' catalytic triad (where 'X' is generally Asp or His) had been recognized in AGOs with slicer activity (Hall, 2005; Nowotny, 2009). Based on analogy to RNase H, a fourth catalytic residue had been suspected, but previous searches for this missing component had focused on the residues corresponding to Arg1045 and Glu1060 (Song et al., 2004; Hall, 2005; Nowotny et al., 2005; Rivas et al., 2005; Yuan et al., 2005), whose conservation and proximity to the catalytic pocket are now explained instead by their roles in stabilizing the plugged-in conformation (Figure 5b). In support of the glutamate finger as the missing catalytic residue that helps to coordinate an active-site metal ion (either directly or through outer-sphere contacts), the putative DEDD catalytic tetrads in the plugged-in conformations of both *Tt*AGO and *Kp*AGO are essentially isosteric with the RNase H DEDD tetrad. Moreover, when we assayed RNAi, only mutation of Glu1013 abrograted RNAi to the extent observed for mutation of Asp1046, a previously identified active-site residue (Figure 5g). Thus, we propose that the glutamate finger constitutes the second residue of a universally conserved RNase H–like DEDX catalytic tetrad at the active site of slicing AGOs.

Our new insights suggest the following model for AGO loading and catalysis. The apo protein in the unplugged conformation binds the siRNA duplex, in part using contacts between the 2-nucleotide overhang of the guide strand and the PAZ domain. As the duplex loads and the 3' end of the guide stand is released from the PAZ domain, the glutamate finger inserts into the active site, thereby completing the DEDX catalytic tetrad to enable cleavage of the passenger strand. After discarding the passenger-strand fragments, the resulting RISC would remain in a plugged-in conformation resembling that of the current structure and be competent to bind and cleave suitably paired target RNAs.

While our manuscript was in review, a structure of human AGO2 (*Hs*AGO2) with RNA of unknown biochemical origin and function was reported (Schirle and Macrae, 2012). The authors noted many contacts to the RNA 5' monophosphate and sugar–phosphate backbone analogous to those of *Kp*AGO. Our inspection of the *Hs*AGO2 structure revealed that *Hs*AGO2 has an extended nucleic acid–binding channel, an N domain positioned to allow unobstructed guide–target pairing, and a plugged-in glutamate finger that completes a DEDH catalytic tetrad (Supplementary Figure 14). These similarities indicate that the *Hs*AGO2 structure (for which conserved residues were mutated to improve diffraction) has some features of active RISC and confirm that *Kp*AGO results provide insights relevant to metazoan AGOs.

In contrast to RNase H, which forms its active site during initial folding, AGO requires a conformational change to form its active site. What might explain this difference between these two related ribonucleases? The constitutive active site of RNase H is well-suited to its role in nonspecifically cleaving RNA–DNA hybrids, whereas proper AGO function requires high specificity. Coupling siRNA-duplex loading (in part through recognition by the PAZ domain) with active-site formation imparts specificity to AGO, thereby preventing it from cleaving any base-paired RNA. After passenger-strand cleavage and removal, activity of the licensed AGO is restricted by its guide RNA. In this way, AGO activity is tightly controlled and spurious endonucleolytic cleavage is prevented. The previous view was that among proteins adopting the RNase H fold, RNase H enzymes were unique in having a catalytic tetrad, whereas the related endonucleases of this protein superfamily (including AGO) were missing the active-site residue corresponding to Glu1013 (Nowotny, 2009). Our findings revising this view imply that some other proteins for which only a catalytic triad has hitherto been identified (e.g., bacterial UvrC DNA repair proteins) might also use the conditional insertion of a “missing” catalytic residue to impart specificity.

Acknowledgements

We thank K. Rajashankar for data processing and phasing, V. Auyeung and D. Shechner for discussions, the Whitehead Genome Technology Core for high-throughput sequencing, and the NE-CAT beamline at the Advanced Photon Source. This work was supported by National Institutes of Health grants AI068776 (D.J.P.) and GM61835 (D.P.B.), a Human Frontier Science Program Long-term Fellowship (K.N.), a fellowship from the Japan Society for the Promotion of Science for Research Abroad (K.N.), and a National Science Foundation graduate research fellowship (D.E.W.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

Accession numbers

The structural coordinates of KpAGO have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) under accession code 4F1N. RNA-sequencing data have been deposited in the Gene Expression Omnibus (http:// www.ncbi.nlm.nih.gov/geo) under accession number GSE37725.

Methods

Protein purification

DNA encoding *K. polysporus* AGO1(Thr 207–Ile 1251) was cloned into a modified pRSFDuet vector (Novagen) containing an amino-terminal Ulp1-cleavable His_6–SUMO tag. Protein was overexpressed in *E. coli* BL21(DE3) Rosetta2 (Novagen). Cell extract was prepared using a French press in buffer A (10 mM phosphate buffer pH 7.3, 1.5 M NaCl, 25 mM imidazole, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride) and cleared by centrifugation. The supernatant was loaded onto a nickel column (GE Healthcare) and then washed with buffer A. The target protein was eluted with a linear gradient of 0.025–1.5 M imidazole. After mixing with Ulp1 protease, the eluted sample was dialysed against buffer B (10 mM phosphate buffer pH 7.3, 500 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol) overnight. The digested protein was loaded onto a nickel column to remove the cleaved His_6–SUMO tag. The flow-through sample was dialysed against buffer C (5 mM phosphate buffer pH 7.3, 10 mM β-mercaptoethanol) and then loaded onto an SP column (GE Healthcare). The protein was eluted with a linear gradient of 0.0–2.0 M NaCl, mixed with ammonium sulphate (2 M final concentration) and then centrifuged. The supernatant was loaded onto a phenyl-Sepharose hydrophobic interaction column (GE Healthcare) in buffer D (10 mM phosphate buffer pH 7.3, 2 M ammonium sulphate, 10 mM β-mercaptoethanol), and the protein was eluted with a linear gradient of 2.0–0.0 M ammonium sulphate. The eluted protein was dialysed against buffer E (300 mM sodium dihydrogen phosphate, 10 mM β-mercaptoethanol) and then loaded onto a MonoQ column (GE Healthcare) in buffer E. The protein was eluted with a linear gradient of 0.0–2.0 M NaCl. The eluted sample was concentrated by ultrafiltration and loaded onto a HiLoad 200 16/60 column (GE Healthcare) in buffer F (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT). Purified KpAGO was concentrated to approximately 40 mg ml^{-1} using ultrafiltration and stored at −80 °C in protein storage buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT).

Structure determination and refinement

Initial crystals of recombinant KpAGO diffracted poorly but could be improved by addition of 1,4-dioxane to the crystallization buffer. Native crystals of KpAGO were obtained at 20 °C by sitting-drop vapour diffusion in 100 mM MIB buffer pH 5.0 (molar ratio, 2 Na-malonate:3 imidazole:3 boric acid), 3% 1,4-dioxane, 19% PEG3350, 12 mM $MnCl_2$ and 3% ethanol. SeMet-substituted crystals were grown at 20 °C by sitting-drop vapour diffusion in 100 mM MIB pH 5.0, 3% 1,4-dioxane, 19% PEG3350, 12 mM $MnCl_2$, 3% ethanol and 9 mM sarcosine. The native and SeMet-substituted crystals of KpAGO were soaked in collection buffer (1.2-fold concentrated reservoir solution) and cryoprotected with 20% glycerol. Both derivative data sets were collected at the Advanced Photon Source NE-CAT beamlines. Data were processed with HKL2000 (Otwinowski and Minor, 1997). Data collection and refinement statistics are listed (Supplementary Table 1). A total of 33 selenium sites were found using peak data with HKL2MAP (Pape and Schneider, 2004) and were used for phase calculation at 4.2 Å resolution with Phaser-EP (McCoy et al., 2007). The initial phases were improved by solvent flattening, electron density histogram and non-crystallographic symmetry averaging with Parrot and DM (Collaborative Computational Project, 1994). The initial model was built manually with

Coot (Jones et al., 1991) and was improved by iterative cycles of refinement with Phenix (Adams et al., 2002). Molecular replacement was performed with MOLREP (Vagin and Teplyakov, 2000) using the SeMet structure as a search model. The final model was improved using the native data processed at 3.2 Å. The Ramachandran plot analysis by PROCHECK (Collaborative Computational Project, 1994) showed 82.0%, 17.3% and 0.8% of the protein residues in the most favourable, additionally allowed and generously allowed regions, respectively, with no residues in disallowed regions. The simulated-annealing omit map was calculated by CNS (Brunger et al., 1998). All figures of structures were generated with PYMOL (DeLano and Lam, 2005).

RNAs

A list of RNA oligonucleotide sequences is provided (Supplementary Table 4). To generate cap-labelled target RNAs, RNA was transcribed *in vitro* with T7 RNA polymerase using DNA oligonucleotide templates. DNase-treated transcripts were purified on a denaturing gel and capped using the ScriptCap m^7G Capping System (CellScript) according to the manufacturer's directions, except that high-specific-activity RNA was prepared by omitting GTP and including 5 µl [α-^{32}P]GTP (6,000 Ci $mmol^{-1}$), and low-specific-activity RNA was prepared by using a 1,500:1 molar ratio of GTP:[α-^{32}P]GTP (6,000 Ci $mmol^{-1}$). Cap-labelled RNA was gel-purified and quantified by scintillation counting, and 10× stocks were prepared in water supplemented with 1 µM DNA carrier oligonucleotide.

5′-phosphorylated guide RNA and its 2′-deoxy-substituted variants were chemically synthesized (IDT) and gel purified. To prepare 5′ end-labelled RNAs, 5′-OH RNAs were chemically synthesized (Dharmacon), deprotected, purified on a denaturing gel, phosphorylated with [γ-^{32}P]ATP (6,000 Ci mmol^{-1}) using T4 polynucleotide kinase (PNK, NEB) and gel purified again. To prepare 3′ end-labelled RNAs, 5′-phosphorylated RNAs lacking the terminal nucleotide (that is, 22-nucleotide variants) were chemically synthesized (IDT), gel purified, extended using cordycepin 5′-[α-^{32}P]triphosphate (5,000 Ci mmol^{-1}) and yeast poly(A) polymerase (USB) and gel purified again.

siRNA duplexes were prepared by annealing synthetic ssRNAs. Complementary RNAs designed to hybridize to generate 21-base-pair duplexes with two-nucleotide 3′ overhangs, were combined (using at least threefold excess unlabelled RNA) in dsRNA annealing buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA) and slow-cooled from 90 °C to room temperature over >2 h. Annealed RNAs were separated from ssRNAs on native 20% polyacrylamide gels, and duplexes were eluted from gel slices in 0.3 M NaCl overnight at 4 °C, ethanol-precipitated and stored in dsRNA storage buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA). RNA was quantified by scintillation counting, and 10× stocks were prepared in dsRNA storage buffer supplemented with 1 μM DNA carrier oligonucleotide.

AGO activity assays

For all biochemical assays, KpAGO was diluted and stored at –20 °C in protein dilution buffer (5 mM Tris-HCl pH 7.5, 100 mM NaCl, 2.5 mM DTT, 50% glycerol). The concentration of KpAGO was determined by absorbance at 280 nm. For the slicing assay in Figure 1c, 1.1 μM KpAGO was pre-incubated with 110 nM guide RNA in 1.1× reaction buffer (1× reaction buffer: 30 mM Tris-HCl pH 7.5, 130 mM KCl, 1.1 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA) for 1 h at 25 °C. To initiate the slicing reaction, 1 μl cap-labelled target RNA (final concentration, 200 nM) was added to 9 μl of the pre-incubated mixture. Reactions were incubated at 30 °C, and 3-μl aliquots were removed at the indicated time and quenched by addition to 12 µl formamide loading buffer (95% formamide, 18 mM EDTA, 0.025% sodium dodecyl sulphate, 0.025% xylene cyanol, 0.025% bromophenol blue). The slicing assay in Figure 2h was conducted similarly except pre-incubation was performed with 110 nM KpAGO and 50 pM guide RNA, and target was subsequently added to a final concentration of 100 pM. The slicing assay in Figure 2f was guided by copurifying RNA and thus did not involve pre-incubation. These reactions contained 1× reaction buffer supplemented with 0.1% Triton X-100, 100 nM KpAGO (or an equal volume of protein dilution buffer) and 100 pM target RNA. Reactions were incubated at 30 °C, and 5 µl aliquots were removed at the indicated time and quenched by addition to 15 µl formamide loading buffer. The passenger-strand cleavage reactions in Figure 3e contained 1× reaction buffer, 100 $\mu g\ ml^{-1}$ Ultrapure BSA (Ambion), 10 nM KpAGO and 50 pM substrate. All other passenger-strand cleavage reactions contained 1× reaction buffer supplemented with 0.1% Triton X-100, 100 nM KpAGO (or an equal volume of protein dilution buffer) and 50 pM substrate. Reactions were incubated at 30 °C, and 5 µl aliquots were removed at the indicated time and quenched by addition to 10–15 µl formamide loading buffer.

To monitor cleavage, RNAs were resolved on denaturing (7.5 M urea) polyacrylamide gels (15% gel for target cleavage using synthetic guide RNA, 20% for target cleavage using copurifying guide RNA or 22.5% for passenger-strand cleavage), and radiolabelled products were visualized by phosphorimaging (Fujifilm BAS-2500) and quantified using Multi Gauge (Fujifilm). For kinetic analyses, at each time point (t) the fraction product was measured as F_P = product/(product + substrate). Data in Figure 3e were fit with a smoothed curve using the cubic spline method implemented in KaleidaGraph.

Analysis of copurifying RNA

To avoid loss of especially small fragments, polynucleotides were extracted without subsequent precipitation. For analysis of pooled crystals, approximately 100 crystals were collected, stored in harvest buffer (100 mM MIB buffer pH 5.0, 3% 1,4-dioxane, 20% PEG 3350, 12 mM $MnCl_2$, 3% ethanol, 6 mM sarcosine, 25% glycerol) and immediately frozen. After thawing, the mixture was diluted with an equal volume of water and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) followed by extraction with chloroform. The aqueous phase was retained and diluted 1:20 in water for use in labelling reactions or used undiluted to prepare sequencing libraries. For analysis of soluble protein, polynucleotides were similarly extracted from 1.5 nmol KpAGO. For analysis of individual crystals, each single crystal was collected, stored in 1 mM EDTA and immediately frozen. After thawing, the mixture was heated at 90 °C for 3 min, chilled on ice for 5 min and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) followed by two extractions with chloroform. The aqueous phase was retained and used undiluted in labelling reactions.

Prior to 5′ labelling, polynucleotides were dephosphorylated in a 20 µl reaction containing 2 µl diluted polynucleotides (or water) and 1× PNK buffer (NEB) in the presence or absence of 2 units of thermosensitive alkaline phosphatase (TSAP, Promega) for 30 min at 37 °C. To inactivate TSAP, the reaction was quenched with 1 µl 220 mM EDTA and incubated at 74 °C for 15 min. 5′ phosphorylation was performed in a 30 µl reaction containing 21 µl heat-inactivated TSAP reaction, 3 units T4 PNK (NEB), 0.04 µl [γ-^{32}P]ATP (8,000 Ci mmol^{-1}), 0.8 µl 10× PNK buffer and 1 µl 240 mM $MgCl_2$ for 1 h at 37 °C. Reactions were quenched with an equal volume of 2× urea loading buffer, and products were resolved on a denaturing 22.5% polyacrylamide gel. For analysis of nuclease sensitivity, 15 µl aliquots were removed from PNK reactions and incubated with 2 µl RNase I (Ambion) or RQ1 RNase-free DNase (Promega) for 30 min at 37 °C before gel analysis.

To monitor preparation of sequencing libraries, trace amounts of synthetic 3′-pCp[5′-^{32}P]-labelled 7- and 23-nucleotide RNA internal standards were added to 2 µl undiluted polynucleotides isolated from soluble or crystalline KpAGO (or a water-only mock control). Dephosphorylation was performed in a 30 µl reaction containing 3 units TSAP (Promega) and 1× PNK buffer (NEB) supplemented with 2 µl manganese-chelating mix (10 mM $MgCl_2$, 10 mM EDTA) for 30 min at 37 °C. To inactivate TSAP, the reaction was quenched with 1.5 µl 240 mM EDTA and incubated at 74 °C for 15 min. RNA was ligated to pre-adenylated adaptor DNA in a 50 µl reaction containing 32 µl heat-inactivated TSAP reaction, 100 pmol adaptor DNA (Grimson et al., 2008), 45 units T4 RNA ligase 1 (NEB), 10% PEG8000 (NEB), 2 µl 10× PNK buffer and 1 µl 390 mM $MgCl_2$ for 2.5 h at room temperature. After phenol extraction and precipitation, 28–50-nucleotide ligation products were gel-purified and 5′ phosphorylated in a 50 µl reaction containing 20 units T4 PNK (NEB) and 1× PNK buffer supplemented with 1 µl [γ-^{32}P]ATP (6,000 Ci mmol^{-1}) for 30 min at 37 °C, followed by a chase with 10 µl cold reaction mixture (1× PNK buffer, 28 units T4 PNK, 6 mM ATP) and incubation for an additional 30 min at 37 °C. After desalting, phenol extraction and precipitation, RNA was ligated to a 5′-adaptor RNA, gel-purified, converted to complementary DNA, amplified 10 cycles (soluble) or 12 cycles (crystalline) and sequenced using the Illumina SBS platform. The library prepared without input polynucleotides did not yield an observable PCR product, indicating minimal contamination from polynucleotides that might copurify with the enzymes used for library construction.

Sequencing reads were filtered by requiring that they contain a perfect match to the first 12 nucleotides of the 3′ adaptor and that every nucleotide up to the beginning of the 3′ adaptor have a Phred+64 quality score of at least ‘ʌ’. After removing the internal-standard reads and trimming away the adaptor sequences, reads representing the small RNAs were collapsed to a non-redundant set of 8–24-nucleotide sequences. To examine the origins of the copurifying RNAs, 15–24-nucleotide sequences were mapped sequentially to the KpAGO expression plasmid, the chloramphenicol-resistance gene found on pRARE2, and the BL21(DE3) genome (Jeong et al., 2009), allowing no mismatches and recovering all hits. (The 15-nucleotide lower bound for mapping was chosen because this was the minimum read length that achieved a <1% genome-mapping rate for random or shuffled small-RNA sequences). Because there were fewer fortuitous matches to the KpAGO expression plasmid, analysis of 12-nucleotide sequences was performed on reads that mapped to the plasmid. For mapping-independent analyses, sequences with <10 reads were not considered.

For analysis of nucleotide composition, information content was calculated by determining the relative frequency of each nucleotide at position *X* compared to the relative frequency at all other positions combined. The selectivity for a given nucleotide *n* at position *X* was calculated using the following equation:

$$S_n = [\Sigma_{i=A,U,G,C}([f(i,X)/f(n,X)]/[f(i,\sim X)/f(n,\sim X)])]^{-1}$$

where $f(i,X)$ is the frequency of nucleotide *i* at position *X* and $f(i,\sim X)$ is the frequency of nucleotide *i* at all other positions. Information content scores were then calculated using the following equation:

$$I_n = S_n \times [\log_2(S_n) + 2]$$

For phasing analysis, the frequency of distances separating 5′-end pairs (i, j) mapping to opposite DNA strands was calculated using the following equation:

$$\mathrm{Frequency}_D = \Sigma_{i,j}\ \min(\mathrm{Reads}_i, \mathrm{Reads}_j)_D$$

where D = (distance between small-RNA 5′ ends) + 1

Yeast manipulations

S. castellii and *S. cerevisiae* were grown at 25 °C and 30 °C, respectively, on standard *S. cerevisiae* plate and liquid media (for example, YPD and SC). Transformations of *S. castellii* were performed as described previously (Drinnenberg et al., 2009). Transformations of *S. cerevisiae* were performed as described (Gietz and Schiestl, 2007). For FACS analyses, strains were inoculated in SC, in either non-inducing (2% glucose) or inducing (1% galactose and 1% raffinose) conditions, and grown overnight. Fresh cultures were then seeded from the overnight cultures, and cells were grown to log phase. Cells were analysed using FACSCalibur (BD Biosciences); data were processed with CellQuest Pro (BD Biosciences) and FlowJo (Tree Star).

Plasmids and strains used and generated in this study are listed (Supplementary Tables 5 and 6). Vectors pRS404CYC1-KpAGO1 and pRS405TEF-KpDCR1 were constructed by insertion of the coding sequencing of the respective *K. polysporus* genes between the *CYC1* or *TEF* promoter and *CYC1* terminator (cloned from p416CYC or p416TEF (Mumberg et al., 1995)) of the appropriate vector (Sikorski and Hieter, 1989) using SpeI and XhoI sites (KpAGO1) or BamHI and XhoI sites (KpDCR1). Vector pRS404CYC1-KpAGO1(207–1251) was constructed similarly, with the insertion of an ‘ATG’ codon upstream of amino acid 207. Vector pRS404CYC1-FLAG$_3$-KpAGO1 was generated by PCR-based insertion of the sequencing encoding the Flag$_3$ epitope downstream of the ‘ATG’ codon of pRS404CYC1-

KpAGO1. Point mutations were introduced by PCR-based mutagenesis to generate vectors encoding mutant Flag-tagged Ago1. pRS402GPD-GFP(S65T) was constructed by insertion of the coding sequence of GFP(S65T) (amplified from pFA6a-GFP(S65T)-kanMX6 (Longtine et al., 1998)) between the *GPD* promoter and *CYC1* terminator (cloned from p416GPD (Mumberg et al., 1995)) of pRS402 (Sikorski and Hieter, 1989) using SpeI and XhoI sites. To reconstitute RNAi in *S. cerevisiae*, GFP(S65T), KpAGO1 and KpDCR1 expression vectors were integrated into W303-1B variants already containing other components of the GFP-silencing system (Drinnenberg et al., 2009), using standard protocols (Gietz and Schiestl, 2007). To generate *S. castellii* strains DPB267 and DPB268 for degradome sequencing, *XRN1* was deleted in DPB005 and DPB007, respectively, using the kanMX6 cassette (Longtine et al., 1998).

Degradome sequencing and analysis

Total RNA was isolated from mid-log phase ($D_{600} \approx 0.6$) cultures of strains DPB267 and DPB268 using the hot-phenol method. Degradome libraries were constructed from 5 μg poly(A)$^+$ RNA essentially as described (Addo-Quaye et al., 2008) and sequenced on the Illumina SBS platform. After removing adaptor sequences and generating each reverse complement, reads representing degradome-cleavage tags were collapsed to a non-redundant set. To analyse tags deriving from Y′-element loci, 20–21-nucleotide sequences were mapped to a consensus *S. castellii* Y′ element as described previously (Drinnenberg et al., 2009), and 49% of reads in the *AGO1* library were randomly sampled (to normalize for higher sequencing yield, Supplementary Figure 2b) and used for subsequent analyses. Mapping data were then used to generate a single-nucleotide-resolution plot of the consensus Y′ element. For phasing analysis, the frequency of distances separating opposite-strand pairs of 5′ ends of 20–21-nucleotide degradome tags (*i*), and 5′ ends of 22–23-nucleotide small RNAs (*j*) was calculated using the following equation:

$$\text{Frequency}_D = \Sigma_{i,j}\ \text{Reads}_i \times \text{Reads}_j/\text{Norm}_i$$

where D = position of 5′ end of degradome tag with respect to 5′ end of small RNA, and Norm_i = number of reads for all small RNAs in which the 5′ end of degradome tag i falls. Fractional frequencies were calculated for each D by dividing Frequency_D by the total number of reads corresponding to degradome-tag 5′ ends that map opposite 22–23-nucleotide small RNAs.

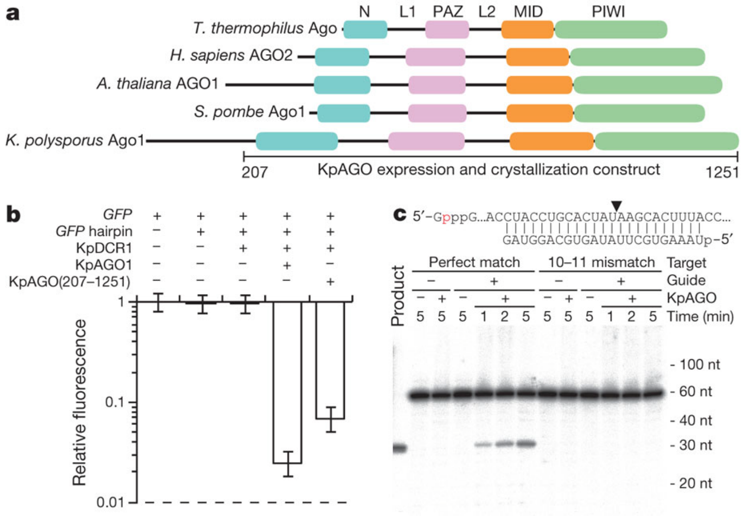


Figure 1. Cleavage activity of budding-yeast AGO
(a) Domain architectures of AGO proteins from *T. thermophilus*, *Homo sapiens*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe* and *K. polysporus*.
(b) RNAi reconstituted in *S. cerevisiae* using *K. polysporus* genes. Median green fluorescent protein (GFP) intensity is plotted as a fraction of GFP-only control. Error bars, quartiles; dashed line, background fluorescence.
(c) Cleavage activity of KpAGO. RNAs labelled at a cap phosphate (red) and matching the guide (either perfectly or with mismatches to positions 10–11) were incubated with/without (+/−) KpAGO that had been pre-incubated in the presence/absence (+/−) of synthetic guide RNA. Product was resolved on a denaturing polyacrylamide gel alongside cap-labelled synthetic product (left) and RNA standards (migration shown on right; nt, nucleotide).

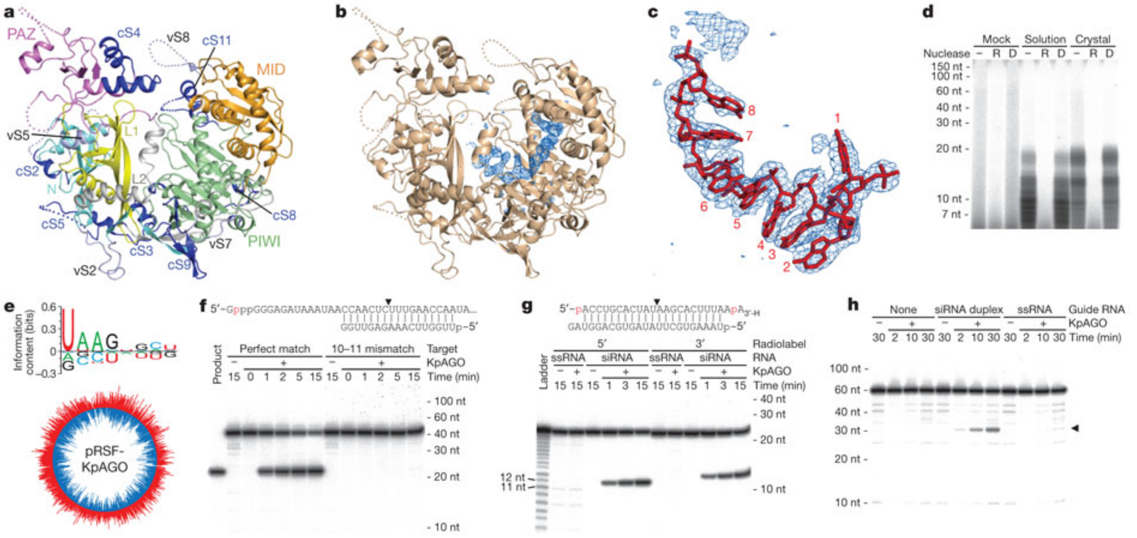


Figure 2. KpAGO architecture and copurifying RNA

(a) KpAGO protein structure, with N (cyan), linker L1 (yellow), PAZ (violet), linker L2 (grey), MID (orange), and PIWI (green) domains in ribbon representation. Constant (cS) and variable (vS) insertion segments, blue and slate, respectively; disordered regions, dotted lines.

(b) $F_o - F_c$ map (blue) contoured at 2.8σ before modelling RNA.

(c) Simulated-annealing omit map (blue) contoured at 3.5σ around final RNA model (red).

(d) Nuclease sensitivity of copurifying nucleic acid. End-labelled polynucleotides extracted from the indicated KpAGO samples were either untreated (−) or incubated with RNase (R) or DNase (D) before analysis on a denaturing gel.

(e) Nucleotide composition and origin of copurifying RNA. Sequences were analysed for enriched or depleted nucleotides (positive or negative bits, respectively) at each of the first eight positions (top). Numbers of sequencing reads mapping along each strand of the KpAGO expression plasmid are indicated (bottom, log scale).

(f) Cleavage activity guided by copurifying RNA. As in Figure 1c, except labelled RNAs were designed to match the indicated copurifying RNA.

(g) Autonomous duplex loading and passenger-strand cleavage. Labelled single-stranded RNA (ssRNA) or siRNA duplex was incubated with or without KpAGO.

(h) Autonomous duplex loading and target cleavage. As in Figure 1c, except KpAGO and RNA concentrations were reduced by 90% and 99.95%, respectively.

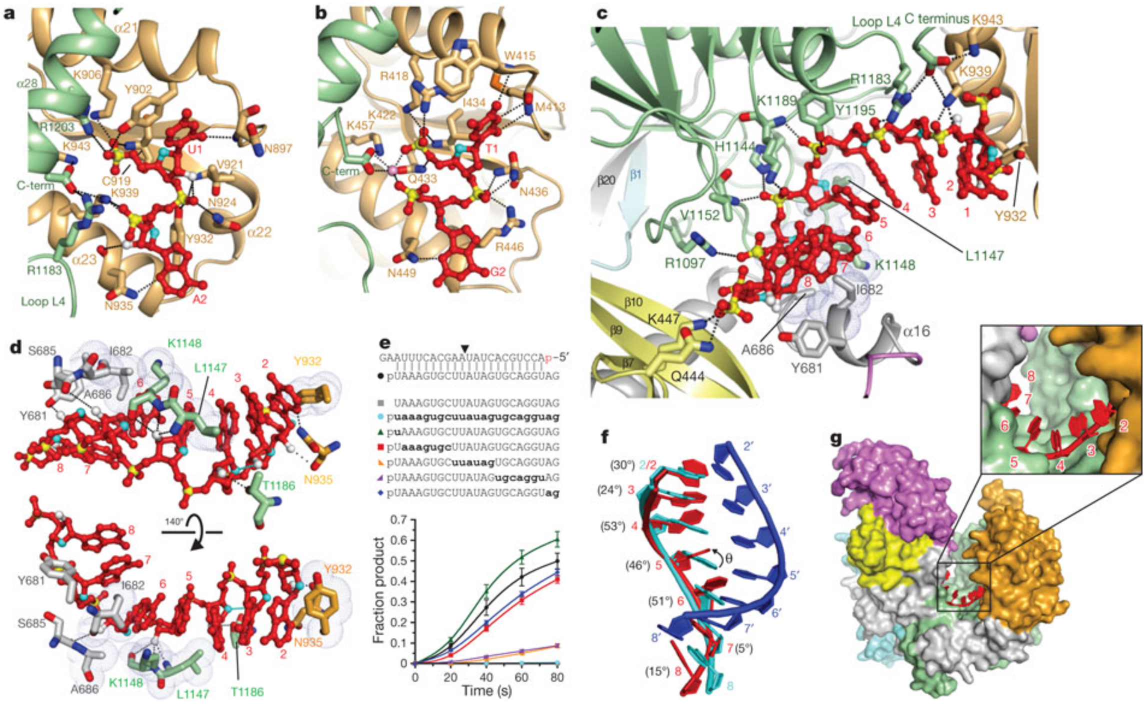


Figure 3. Organization of the guide RNA

(a, b) The 5′-nucleotide-binding pockets of KpAGO (a) and TtAGO (Wang et al., 2008b) (b). Colours, as in Figure 2a; protein, ribbon representation; highlighted residues and RNA, stick representation; O2′, O4′ and phosphate, white, cyan and yellow, respectively; hydrogen bonds, dotted lines.

(c, d) Interactions involving bases and either phosphates (c) or 2′-OH groups (d) of the seed region. Intermolecular (black) and intramolecular (blue) hydrogen bonds, dotted lines; hydrophobic interactions, van der Waals radii.

(e) Effects of guide-strand modifications on duplex loading and passenger-strand cleavage. KpAGO was incubated with siRNA duplexes with the indicated guide strands; p, 5′ monophosphate; upper case, 2′-OH; lower case bold, 2′-deoxy. The fraction of labelled passenger strand cleaved is plotted (average of three independent replicates; error bars, standard deviations; points connected by smooth curves).

(f) Superposition of guide-RNA nucleotides 2–8 (red) on A-form RNA (cyan and blue). Dihedral angles (θ) between guide-RNA bases and those of A-form RNA are in parentheses.

(g) Solvent-exposed seed nucleotides (red). KpAGO surface is rendered, domains coloured as in Figure 2a.

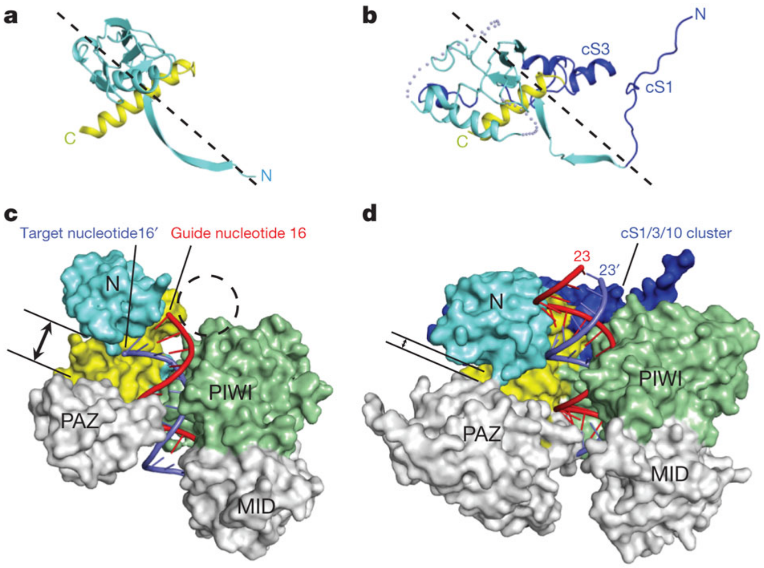


Figure 4. An extended, potentially unobstructed nucleic-acid-binding channel in KpAGO
(a, b) Position of N domain (cyan) relative to L1-linker domain (yellow) in TtAGO (Wang et al., 2009) (a) and KpAGO (b). Domains are oriented based on their N-terminal beta strands (dashed line connects strand termini). Colours, as in Figure 2a.
(c, d) Channels of TtAGO ternary complex (Wang et al., 2009 (c) and KpAGO with modelled A-form duplex (d). Protein surfaces are rendered, highlighting distances between the N and PAZ domains (parallel lines) and the cS1/3/10 cluster (blue), which fills a cavity (dashed circle).

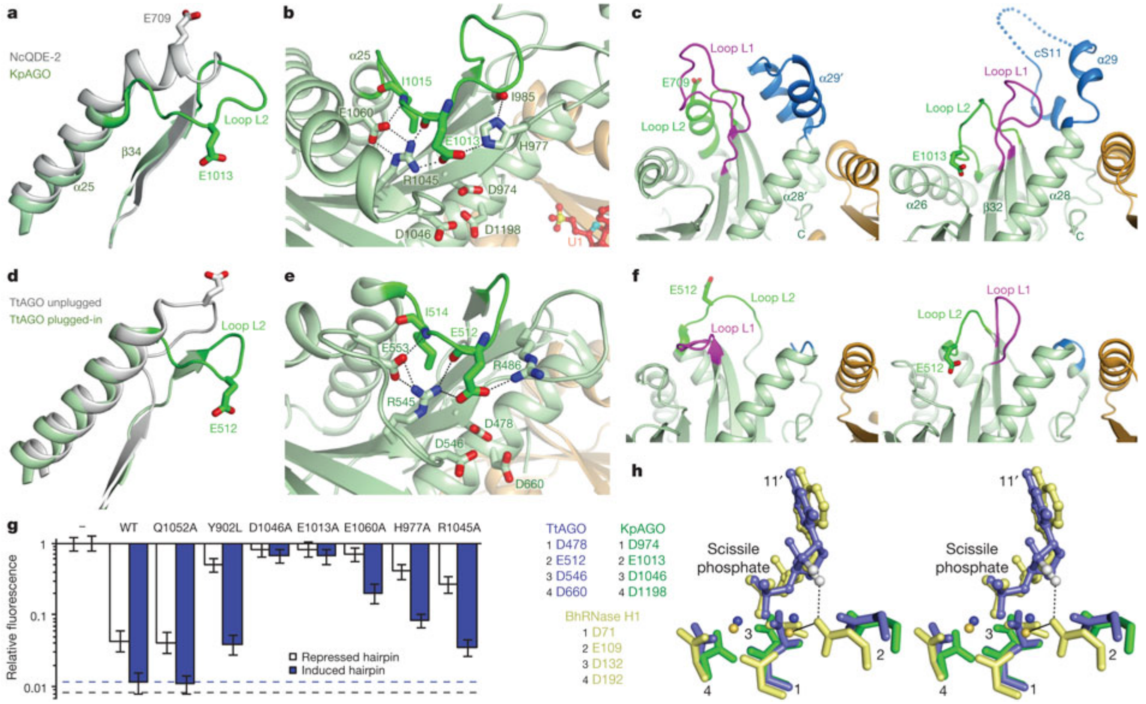


Figure 5. A plugged-in glutamate finger at the active site
(a) Superposition of α25 and β34 of KpAGO (green) on counterparts of NcQDE-2 MID-PIWI lobe (grey), highlighting the extended loop L2 (dark green).
(b) Hydrogen-bond network stabilizing the plugged-in loop L2. Loop L2, dark green; otherwise, as in Figure 3a.
(c) Closed (left) and open (right) configurations of the loop L1 gate (purple) in NcQDE-2 MID-PIWI lobe and KpAGO, respectively. cS11, blue; otherwise, as in panel b.
(d) Superposition of the region flanking loop L2 in the unplugged (grey) and plugged-in (green) conformations of TtAGO, depicted as in panel a.
(e) Hydrogen-bond network stabilizing the plugged-in loop L2 in TtAGO, depicted as in panel b.
(f) Closed and open configurations of the loop L1 gate in the unplugged and plugged-in conformations of TtAGO, respectively, depicted as in panel c.
(g) RNAi reconstituted in *S. cerevisiae* using wild-type (WT) *K. polysporus AGO1* or genes with the indicated substitutions. Silencing was monitored under either permissive (induced hairpin, blue bars) or stringent (repressed hairpin, open bars) conditions. Q1052 and Y902, conserved residues insensitive and sensitive to substitution, respectively (Boland et al., 2011), were included as controls. Dashed lines (blue and black), background fluorescence (permissive and stringent conditions, respectively); otherwise, as in Figure 1b.
(h) Stereoview of KpAGO catalytic residues (green) superpositioned with catalytic residues, divalent cations, scissile phosphate and adjacent nucleoside in TtAGO (blue) and *B. halodurans* RNase H1 (BhRNase H1, yellow) ternary complexes.

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol *18*, 758-762.

Ameres, S.L., Martinez, J., and Schroeder, R. (2007). Molecular basis for target RNA recognition and cleavage by human RISC. Cell *130*, 101-112.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep *11*, 522-527.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr Sect D-Biol Crystallogr *54*, 905-921.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Chiu, Y.L., and Rana, T.M. (2003). siRNA function in RNAi: a chemical modification analysis. RNA *9*, 1034-1048.

Collaborative Computational Project (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

DeLano, W.L., and Lam, J.W. (2005). PyMOL: A communications tool for computational models. Abstracts of Papers of the American Chemical Society *230*.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544-550.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Frazao, C., McVey, C.E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C.M., and Carrondo, M.A. (2006). Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. Nature *443*, 110-114.

German, M.A., Pillay, M., Jeong, D.H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R., *et al.* (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol *26*, 941-946.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc *2*, 31-34.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol *11*, 599-606.

Hall, T.M. (2005). Structure and function of argonaute proteins. Structure *13*, 1403-1408.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.H., Couloux, A., Lee, S.W., Yoon, S.H., Cattolico, L., *et al.* (2009). Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). J Mol Biol *394*, 644-652.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A *47 (Pt 2)*, 110-119.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E.V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct *4*, 29.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. (2004). MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. EMBO J *23*, 3356-3364.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev *18*, 975-980.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Curr Biol *15*, 2149-2155.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119-122.

Nowotny, M. (2009). Retroviral integrase superfamily: the structural perspective. EMBO Rep *10*, 144-151.

Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell *121*, 1005-1016.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, C.W. Carter, and R.M. Sweet, eds. (San Diego, Academic Press), pp. 307-326.

Pan, Y., and MacKerell, A.D., Jr. (2003). Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. Nucleic Acids Res *31*, 7131-7140.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. J Appl Cryst *37*, 843-844.

Parker, J.S. (2010). How to slice: snapshots of Argonaute in action. Silence *1*, 3.

Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell *33*, 204-214.

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Schirle, N.T., and Macrae, I.J. (2012). The Crystal Structure of Human Argonaute2. Science *336*, 1037-1040.

Schwarz, D.S., Hutvagner, G., Haley, B., and Zamore, P.D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell *10*, 537-548.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev *19*, 517-529.

Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr *56*, 1622-1624.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008a). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754-761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008b). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Weinberg, D.E., Nakanishi, K., Patel, D.J., and Bartel, D.P. (2011). The inside-out mechanism of Dicers from budding yeasts. Cell *146*, 262-276.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Supplementary Table 1. X-ray statistics of data collection and refinement

	Native	Se-SAD
Data collection		
Space group	$P3_2$	$P3_2$
Cell dimensions		
a, b, c (Å)	171.55, 171.55, 83.18	168.92, 168.92, 84.25
α, β, γ (º)	90, 90, 120	90, 90, 120
Wavelength (Å)	0.97949	0.97916
Resolution (Å)	50.0–3.2 (3.31–3.20)	50.0–4.15 (4.30–4.15)
Measured reflections	262,889	189,051
Unique reflections	45,536	40,658
R_{merge}	0.186 (0.736)	0.162 (0.469)
$I/\sigma I$	12.8 (3.0)	10.3 (3.6)
Completeness (%)	100 (99.7)	99.7 (99.8)
Redundancy	5.8 (5.6)	4.6 (4.7)
Refinement		
Resolution (Å)	38.12–3.19	
Number of reflections	45,435	
$R_{\mathrm{work}}/R_{\mathrm{free}}$	16.83/21.62	
Number of atoms		
Protein	14,034	
RNA	358	
Water	34	
B factors (Å^2)		
Protein	52.47	
RNA	42.6	
Water	28.6	
R.m.s deviations		
Bond lengths (Å)	0.010	
Bond angles (º)	1.347	

Highest resolution shells are shown in parentheses.

Supplementary Table 2. Analysis of small-RNA libraries

Filtering reads	Soluble		Crystalline	
Total reads	27,117,105		26,599,870	
No adaptor sequence	2,489,636	(9.2)	2,246,179	(8.4)
Standard used in library construction	1,230	(0.0)	2,322	(0.0)
Small-RNA reads	24,626,239	(90.8)	24,351,369	(91.5)
Failed quality filter	3,875,605	(15.7)	3,890,220	(16.0)
Passed quality filter	20,750,634	(84.3)	20,461,149	(84.0)
15–24 nt for mapping	1,035,603	(5.0)	1,676,927	(8.2)
8–24 nt with ≥10 reads for scatterplot	10,659,771	(51.4)	11,428,129	(55.9)
Unique sequences for scatterplot	63,144		72,024	

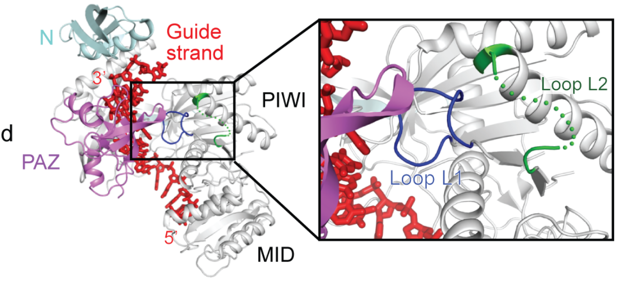
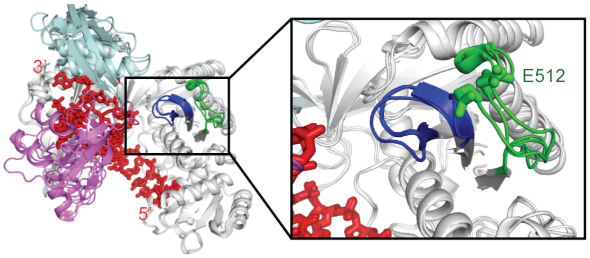
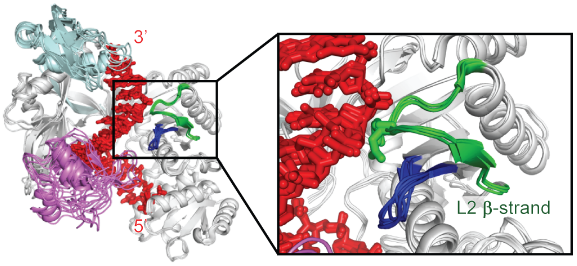
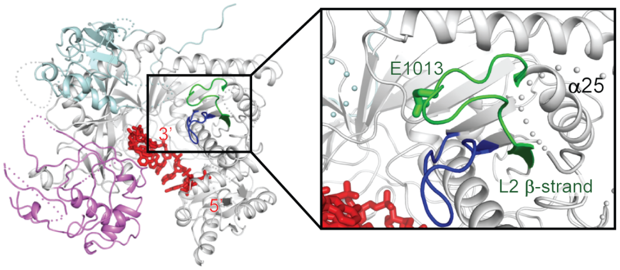
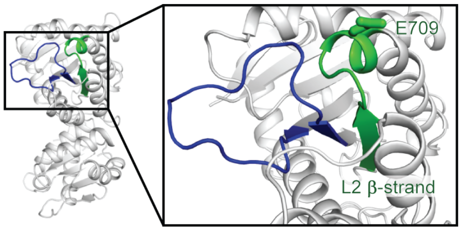
Percentages in parentheses are relative to the parent category. The scatterplot is Supplementary Figure 5c.

Mapping reads	Soluble				Crystalline			
	Reads		Sequences		Reads		Sequences	
pRSF-KpAGO	846,852	(81.8)	30,884	(23.1)	1,309,717	(78.1)	32,677	(16.9)
Cam^{R} gene	2,877	(0.3)	684	(0.5)	5,311	(0.3)	891	(0.5)
BL21(DE3)	81,526	(7.9)	39,295	(29.3)	158,596	(9.5)	67,958	(35.2)
Unmapped	104,348	(10.1)	63,086	(47.1)	203,303	(12.1)	91,794	(47.5)
Total	1,035,603	(100)	133,949	(100)	1,676,927	(100)	193,320	(100)

Percentages in parentheses are relative to the total number of 15–24-nucleotide reads or sequences passing the quality filter.

Supplementary Table 3. Comparison of the crystal structures of KpAGO, TtAGO and NcQDE-2 MID-PIWI lobe

Structures are drawn in ribbon representation, with the N (cyan) and PAZ (magenta) domains highlighted and loops L1 and L2 coloured in blue and green, respectively. Guide strands (red) are shown in stick representation. For clarity, the target strands are not shown in the TtAGO ternary complexes.

Structure	PDB code	Guide 3′ end	Conformation of loop L2	
TtAGO binary complex with a guide DNA	3DLH	PAZ	Unplugged	
TtAGO ternary complex with guide DNA and target RNA	3F73 3HO1 3HXM	PAZ	Unplugged	
TtAGO ternary complex with guide DNA and target RNA	3HJF 3HK2 3HM9 3HVR	Released	Plugged	
KpAGO binary complex with guide RNA	Current structure	Released	Plugged	
NcQDE-2 MID-PIWI lobe in apo form	2YHA	–	Unplugged	

Supplementary Table 4. Plasmids used and generated in this study

Plasmid	**Description**
pRSF-KpAGO	*E. coli* expression plasmid, *K. polysporus AGO1*(207–1251)
pRS402GPD-GFP(S65T)	Integrating plasmid, GFP(S65T) under *GPD* promoter
pRS405TEF-KpDCR1	Integrating plasmid, *K. polysporus DCR1* under *TEF* promoter
pRS404CYC-KpAGO1	Integrating plasmid, *K. polysporus AGO1* under *CYC1* promoter
pRS404CYC-KpAGO1(207–1251)	Integrating plasmid, *K. polysporus AGO1*(207–1251) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1* under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(Y902L)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(Y902L) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(H977A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(H977A) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(E1013A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(E1013A) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(R1045A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(R1045A) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(D1046A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(D1046A) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(Q1052A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(Q1052A) under *CYC1* promoter
pRS404CYC-FLAG$_3$-KpAGO1(E1060A)	Integrating plasmid, FLAG$_3$-tagged *K. polysporus AGO1*(E1060A) under *CYC1* promoter

Supplementary Table 5. Sequences of RNA oligonucleotides used in this study

Guide RNA[1]
UAAAGUGCUUAUAGUGCAGGUAG

Passenger RNA
ACCUGCACUAUAAGCACUUUAAG

Guide RNA for 3′ labelling
UAAAGUGCUUAUAGUGCAGGUA

Passenger RNA for 3′ labelling
ACCUGCACUAUAAGCACUUUAA

Perfect-match target for synthetic guide RNA
GGGAGAAACAAAAAUACCUACCUGCACUAUAAGCACUUUACCAUCUCAAACUUACUCAGA

10–11-mismatch target for synthetic guide RNA[2]
GGGAGAAACAAAAAUACCUACCUGCACUA**AU**AGCACUUUACCAUCUCAAACUUACUCAGA

Expected cleavage product for synthetic guide RNA
GGGAGAAACAAAAAUACCUACCUGCACUAU

Perfect-match target for copurifying 17-nucleotide RNA
GGGAGAUAAAUAACCAACUCUUUGAACCAAUAUAUGAAGC

10–11-mismatch target for copurifying 17-nucleotide RNA[2]
GGGAGAUAAAUAACCAACU**GA**UUGAACCAAUAUAUGAAGC

Expected cleavage product for copurifying 17-nucleotide RNA
GGGAGAUAAAUAACCAACUC

---

[1] 2′-deoxy–substituted guide RNAs used in passenger-strand cleavage studies are depicted in Figure 3e.

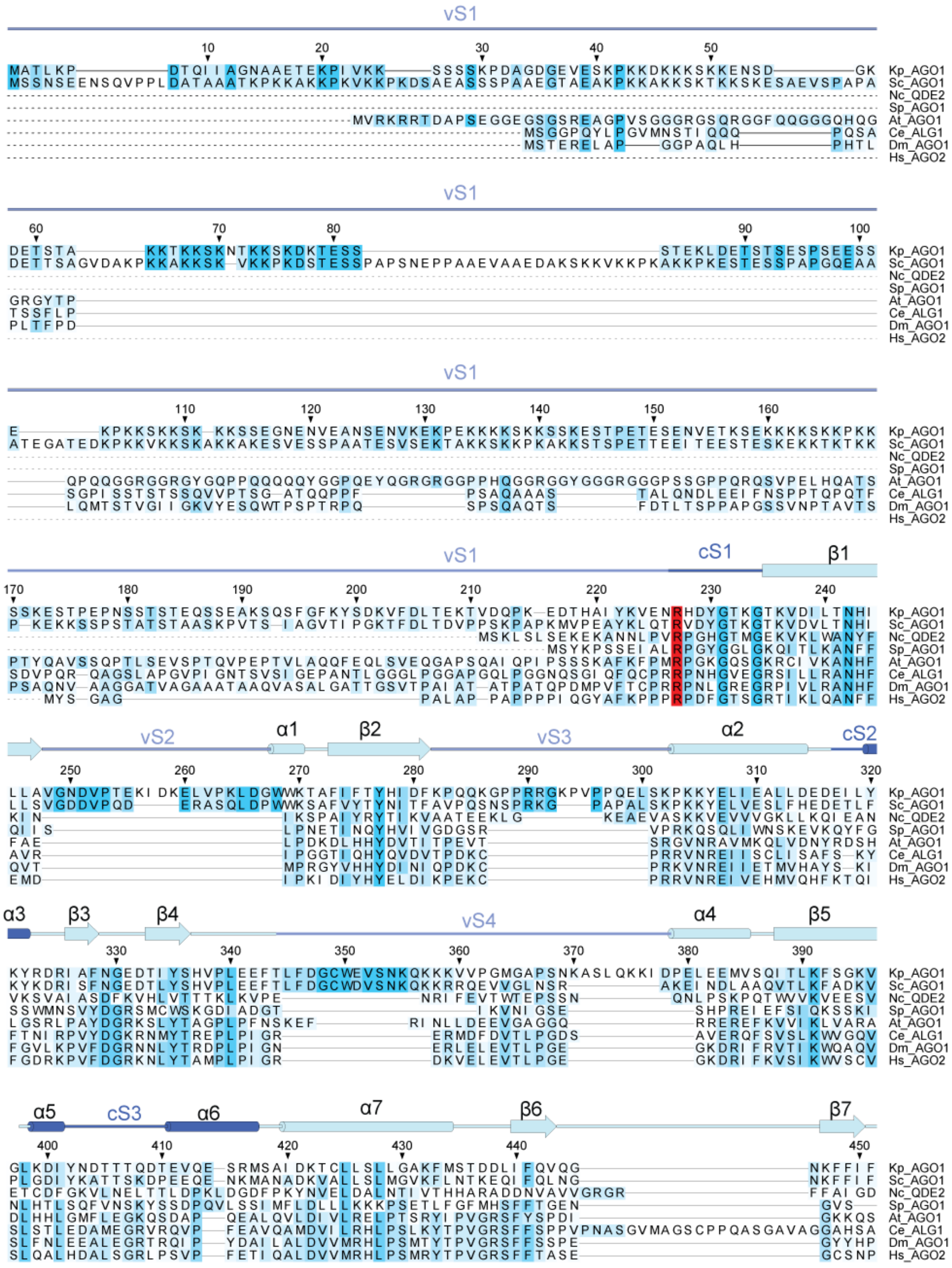
[2] Bases that form mismatches with the guide RNA are bold and underlined.

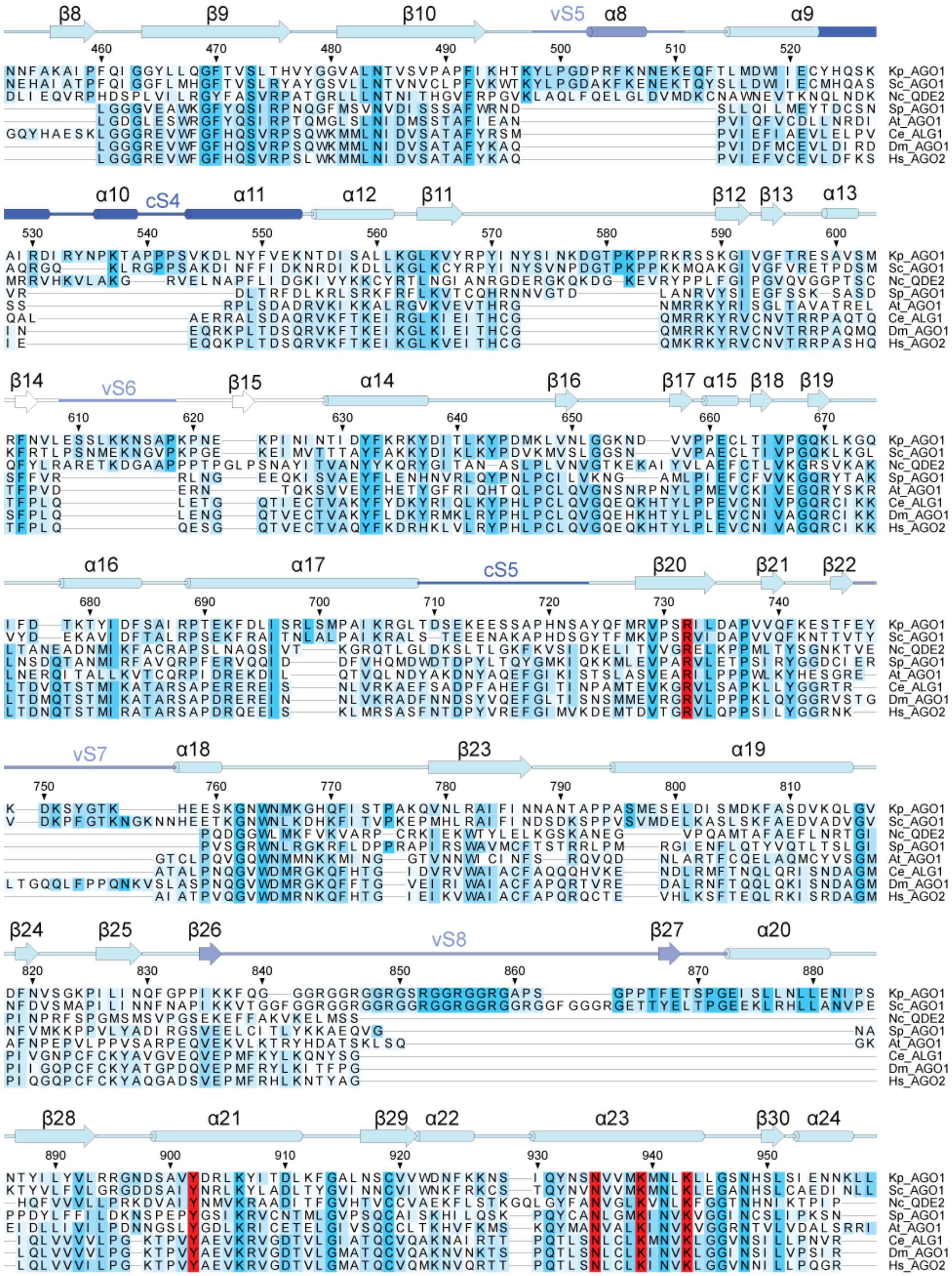
Supplementary Table 6. Yeast strains used and generated in this study

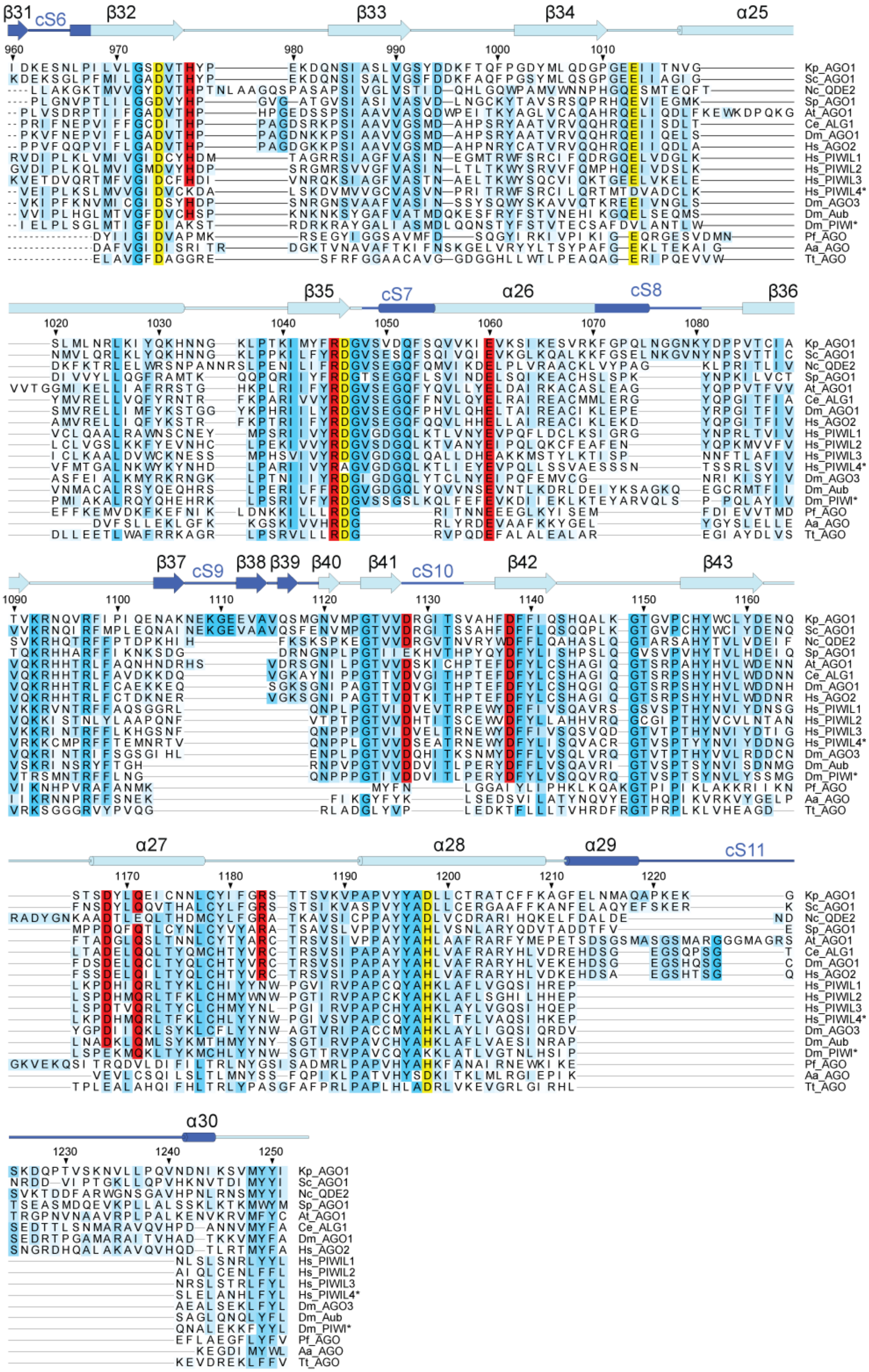
Strain	Genotype	Species	Reference
DPB005	*MATα ho*Δ *ura3-1*	*S. castellii*	*A*
DPB007	*MATα ho*Δ *ura3-1 ago1*Δ::Hyg	*S. castellii*	*A*
DPB267	*MATα ho*Δ *ura3-1 xrn1*Δ::KanMX6	*S. castellii*	This study
DPB268	*MATα ho*Δ *ura3-1 ago1*Δ::Hyg *xrn1*Δ::KanMX6	*S. castellii*	This study
L4718	*MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*	*S. cerevisiae*	*B*
DPB900	*MATα leu2-3,112 trp1-1 can1-100 ura3-1 ADE2::pGPD(empty) his3-11,15*	*S. cerevisiae*	This study
DPB249	*MATα leu2-3,112 trp1-1 can1-100 ura3::GFP(S65T)-KanMX6 ade2-1 his3-11,15*	*S. cerevisiae*	*A*
DPB914	*MATα leu2-3,112 trp1-1 can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) his3-11,15*	*S. cerevisiae*	This study
DPB915	*MATα leu2-3,112 trp1-1 can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB901	*MATα LEU2::pTEF-KpDCR1 trp1-1 can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB916	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-KpAGO1 can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB917	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-KpAGO1(207–1251) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB906	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1 can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB907	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(Y902L) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB908	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(H977A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB909	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(E1013A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB910	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(R1045A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB911	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(D1046A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB912	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(Q1052A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study
DPB913	*MATα LEU2::pTEF-KpDCR1 TRP1::pCYC-FLAG*$_3$*-KpAGO1(E1060A) can1-100 ura3::GFP(S65T)-KanMX6 ADE2::pGPD-GFP(S65T) HIS3::pGAL1-strongSC_GFP*	*S. cerevisiae*	This study

A. Drinnenberg, et al. RNAi in budding yeast. *Science* **326**, 544-50 (2009).

B. Thomas & Rothstein. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619-30 (1989).


vS1
cS1
β1
vS2
α1
β2
vS3
α2
cS2
α3
β3
β4
vS4
α4
β5
α5
cS3
α6
α7
β6
β7
Kp_AGO1
Sc_AGO1
Nc_QDE2
Sp_AGO1
At_AGO1
Ce_ALG1
Dm_AGO1
Hs_AGO2

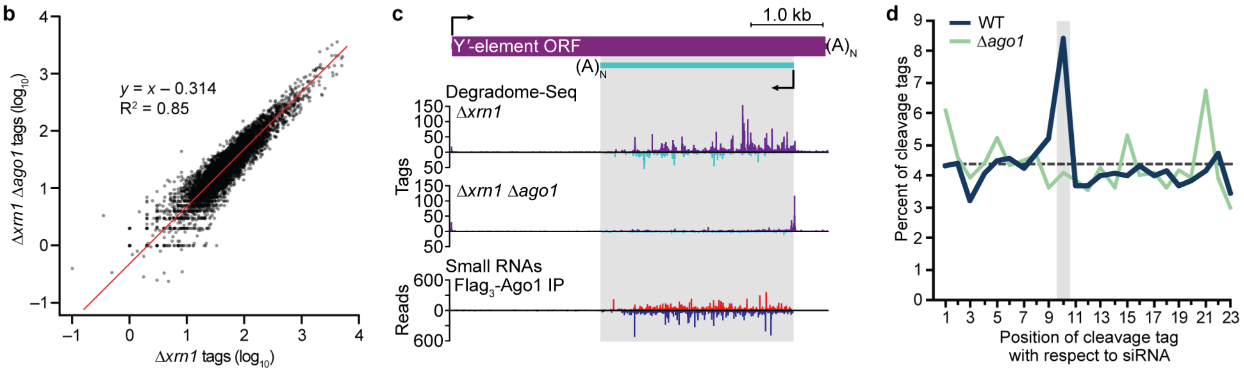

β8 β9 β10 vS5 α8 α9
460 470 480 490 500 510 520
NNFAKAIPFQIGGYLLQGFTVSLTHVYGGVALNTVSVPAPFIKHTKYLPGDPRFKNNEKEQFTLMDWIIECYHQSK Kp_AGO1
NEHAIATPFQIGGFLMHGFTVSLRYAYGSVLLNTVNVCLPFVKWTKYLPGDAKFKENEKTQYSLLDWIIECMHQAS Sc_AGO1
DLIEQVRPHDSPLVILRGYFASVRPATGRLLLNTNITHGVFRPGVKLAQLFQELGLDVMDKCNAWNEVTKNQLNDK Nc_QDE2
-LGGGVEAWKGFYQSIRPNQGFMSVNVDISSSAFWRND-SLLQILMEYTDCSN Sp_AGO1
-LGDGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEAN-PVIQFVCDLLNRDI At_AGO1
GQYHAESKLGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYRSM-PVIEFIAEVLELPV Ce_ALG1
-LGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYKAQ-PVIDFMCEVLDIRD Dm_AGO1
-LGGGREVWFGFHQSVRPSLWKMMLNIDVSATAFYKAQ-PVIEFVCEVLDFKS Hs_AGO2
α10 cS4 α11 α12 β11 β12 β13 α13
530 540 550 560 570 580 590 600
AIRDIRYNPKTAPPPSVKDLNYFVEKNTDISALLKGLKVYRPYINYSINKDGTPKPPRKRSSKGIVGFTRESAVSM Kp_AGO1
AQRGQ-KLRGPPSAKDINFFIDKNRDIKDLLKGLKCYRPYINYSVNPDGTPKPPKKMQAKGIVGFVRETPDSM Sc_AGO1
MRRVHKVLAKG-RVELNAPFLIDGKIVYKKCYRTLNGIANRGDERGKQKDG-KEVRYPPLFGIPGVQVGGPTSC Nc_QDE2
VR-DLTRFDLKRLSRKFRFLKVTCQHRNNVGTD-LANRVYSIEGFSSK-SASD Sp_AGO1
SS-RPLSDADRVKIKKALRGVKVEVTHRG-NMRRKYRISGLTAVATREL At_AGO1
QAL-AERRALSDAQRVKFTKEIRGLKIEITHCG-QMRRKYRVCNVTRRPAQTQ Ce_ALG1
IN-EQRKPLTDSQRVKFTKEIKGLKIEITHCG-QMRRKYRVCNVTRRPAQMQ Dm_AGO1
IE-EQQKPLTDSQRVKFTKEIKGLKVEITHCG-QMKRKYRVCNVTRRPASHQ Hs_AGO2
β14 vS6 β15 α14 β16 β17 α15 β18 β19
610 620 630 640 650 660 670
RFNVLESSLKKNSAPKPNE-KPININTIDYFKRKYDITLKYPDMKLVNLGGKND-VVPPECLTIVPGQKLKGQ Kp_AGO1
KFRTLPSNMEKNGVPKPGE-KEIMVTTTAYFAKKYDIKLKYPDVKMVSLGGSN-VVPAECLTIVPGQKLKGL Sc_AGO1
QFYLRARETKDGAAPPPTPGLPSNAYITVANYYKQRYGITAN-ASLPLVNVGTKEKAIYVLAEFCTLVKGRSVKAK Nc_QDE2
SFFVR-RLNG-EEQKISVAEYFLENHNVRLQYPNLPCILVKNG-AMLPIEFCFVVKGQRYTAK Sp_AGO1
TFPVD-ERN-TQKSVVEYFHETYGFRIQHTQLPCLQVGNSNRPNYLPMEVCKIVEGQRYSKR At_AGO1
TFPLQ-LETG-QTIECTVAKYFYDKYRIQLKYPHLPCLQVGQEQKHTYLPPEVCNIVPGQRCIKK Ce_ALG1
SFPLQ-LENG-QTVECTVAKYFLDKYRMKLRYPHLPCLQVGQEHKHTYLPLEVCNIVAGQRCIKK Dm_AGO1
TFPLQ-QESG-QTVECTVAQYFKDRHKLVLRYPHLPCLQVGQEQKHTYLPLEVCNIVAGQRCIKK Hs_AGO2
α16 α17 cS5 β20 β21 β22
680 690 700 710 720 730 740
IFD-TKTYIDFSAIRPTEKFDLISRLSMPAIKRGLTDSEKEESSAPHNSAYQFMRVPSRILDAPVVQFKESTFEY Kp_AGO1
VYD-EKAVIDFTALRPSEKFRAITNLALPAIKRALS-TEEENAKAPHDSGYTFMKVPSRVIDAPVVQFKNTTVTY Sc_AGO1
LTANEADNMIKFACRAPSLNAQSIVT-KGRQTLGLDKSLTLGKFKVSIDKELITVVGRELKPPMLTYSGNKTVE Nc_QDE2
LNSDQTANMIRFAVQRPFERVQQID-DFVHQMDWDTDPYLTQYGMKIQKKMLEVPARVLETPSIRYGGDCIER Sp_AGO1
LNERQITALLKVTCQRPIDREKDIL-QTVQLNDYAKDNYAQEFGIKISTSLASVEARILPPPWLKYHESGRE- At_AGO1
LTDVQTSTMIKATARSAPEREREIS-NLVRKAEFSADPFAHEFGITINPAMTEVKGRVLSAPKLLYGGRTR- Ce_ALG1
LTDMQTSTMIKATARSAPDREREIN-NLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQYGGRVSTG Dm_AGO1
LTDNQTSTMIRATARSAPDRQEEIS-KLMRSASFNTDPYVREFGIMVKDEMTDVTGRVLQPPSILYGGRNK- Hs_AGO2
vS7 α18 β23 α19
750 760 770 780 790 800 810
K-DKSYGTK-HEESKGNWNMKGHQFISTPAKQVNLRAIFINNANTAPPASMESELDISMDKFASDVKQLGV Kp_AGO1
V-DKPFGTKNGKNNHEETKGNWNLKDHKFITVPKEPMHLRAIFINDSDKSPPVSVMDELKASLSKFAEDVADVGV Sc_AGO1
-PQDGGWLMKFVKVARP-CRKIEKWTYLELKGSKANEG-VPQAMTAFAEFLNRTGI Nc_QDE2
-PVSGRWNLRGKRFLDPPRAPIRSWAVMCFTSTRRLPM-RGIENFLQTYVQTLTSLGI Sp_AGO1
-GTCLPQVGQWNMMNKKMING-GTVNNWICINFS-RQVQD-NLARTFCQELAQMCYVSGM At_AGO1
-ATALPNQGVWDMRGKQFHTG-IDVRVWAIACFAQQQHVKE-NDLRMFTNQLQRISNDAGM Ce_ALG1
LTGQQLFPPQNKVSLASPNQGVWDMRGKQFFTG-VEIRIWAIACFAPQRTVRE-DALRNFTQQLQKISNDAGM Dm_AGO1
-AIATPVQGVWDMRNKQFHTG-IEIKVWAIACFAPQRQCTE-VHLKSFTEQLRKISRDAGM Hs_AGO2
β24 β25 β26 vS8 β27 α20
820 830 840 850 860 870 880
DFNVSGKPILINQFGPPIKKFQG-GGRGGRGGRGSRGGRGGRGAPS-GPPTFETSPGEISLLNLLENIPS Kp_AGO1
NFDVSMAPILINNFNAPIKKVTGGFGGRGGRGGRGGRGGRGGRGGRGGFGGGRGETTYELTPGEEKLRHLLANVPE Sc_AGO1
PINPRFSPGMSMSVPGSEKEFFAKVKELMSS- Nc_QDE2
NFVMKKPPVLYADIRGSVEELCITLYKKAEQVG-NA Sp_AGO1
AFNPEPVLPPVSARPEQVEKVLKTRYHDATSKLSQ-GK At_AGO1
PIVGNPCFCKYAVGVEQVEPMFKYLKQNYSG- Ce_ALG1
PIIGQPCFCKYATGPDQVEPMFRYLKITFPG- Dm_AGO1
PIQGQPCFCKYAQGADSVEPMFRHLKNTYAG- Hs_AGO2
β28 α21 β29 α22 α23 β30 α24
890 900 910 920 930 940 950
NTYILYVLRRGNDSAVYDRLKYITDLKFGALNSCVVWDNFKKNS-IQYNSNVVMKMNLKLLGSNHSLSIENNKLL Kp_AGO1
KTYVLFVLGRGDDSAIYNRLKYLADLTYGVINNCVIWNKFRKCS-TQYNVNVVMKMNLKLEGANHSLCAEDINLL Sc_AGO1
-HQFVVVLLPRKDVAIYNMVKRAADITFGVHTVCCVAEKFLSTKGQLGYFANVGLKVNLKFGGTNHNIKTPIP- Nc_QDE2
PPDYLFFILLDKNSPEPYGSIKRVCNTMLGVPSQCAISKHILQSK-PQYCANLGMKINVKVGGINCSLIPKSN- Sp_AGO1
EIDLLIVILPDNNGSLYGDLKRICETELGIVSQCCLTKHVFKMS-KQYMANVALKINVKVGGRNTVLVDALSRRI At_AGO1
-IQLVVVVLPG-KTPVYAEVKRVGDTVLGIATQCVQAKNAIRTT-PQTLSNLCLKMNVKLGGVNSILLPNVR- Ce_ALG1
-LQLVVVVLPG-KTPVYAEVKRVGDTVLGMATQCVQAKNVNKTS-PQTLSNLCLKINVKLGGINSILVPSIR- Dm_AGO1
-LQLVVVILPG-KTPVYAEVKRVGDTVLGMATQCVQMKNVQRTT-PQTLSNLCLKINVKLGGVNNILLPQGR- Hs_AGO2


β31
cS6
β32
β33
β34
α25
960
970
980
990
1000
1010
β35
cS7
α26
cS8
β36
1020
1030
1040
1050
1060
1070
1080
β37
cS9
β38
β39
β40
β41
cS10
β42
β43
1090
1100
1110
1120
1130
1140
1150
1160
α27
α28
α29
cS11
1170
1180
1190
1200
1210
1220
α30
1230
1240
1250
Kp_AGO1
Sc_AGO1
Nc_QDE2
Sp_AGO1
At_AGO1
Ce_ALG1
Dm_AGO1
Hs_AGO2
Hs_PIWIL1
Hs_PIWIL2
Hs_PIWIL3
Hs_PIWIL4*
Dm_AGO3
Dm_Aub
Dm_PIWI*
Pf_AGO
Aa_AGO
Tt_AGO

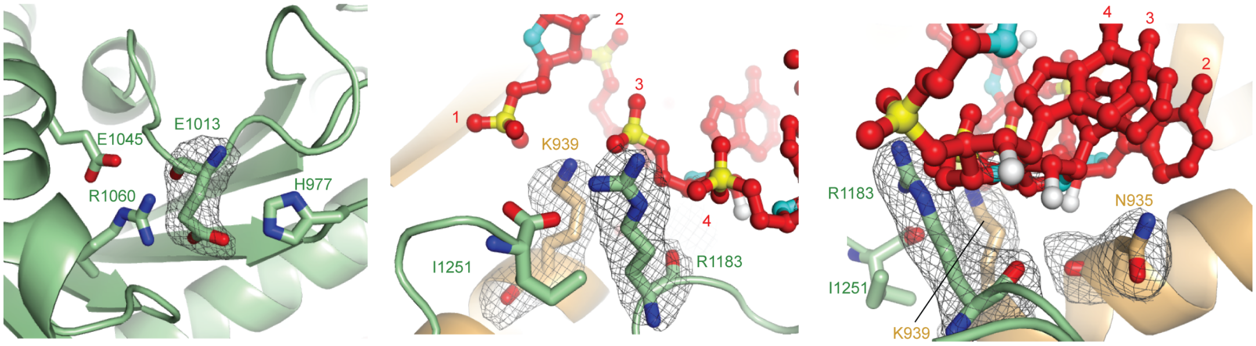
Supplementary Figure 1. Sequence alignment of the AGO family
Protein sequences were aligned across their length (eukaryotic AGOs) or across their PIWI domain (prokaryotic AGOs and eukaryotic PIWI-clade proteins). Highlighted are the four residues of the catalytic tetrad (yellow), other residues analysed in this work (red), and additional well-conserved amino acids (blue; intensity indicates degree of conservation). Residue numbers and the secondary structure of KpAGO are indicated above the alignment. Kp, *K. polysporus*; Sc, *S. castellii*; Nc, *N. crassa*; Sp, *S. pombe*; At, *A. thaliana*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *H. sapiens*; Pf, *P. furiosus;* Aa, *A. aeolicus*; and Tt, *T. thermophilus*. PIWI-clade proteins for which the residue corresponding to Glu 1013 is not conserved are indicated (*); each of these also contains a non-conserved amino acid at the position of one of the previously proposed active-site residues.

a

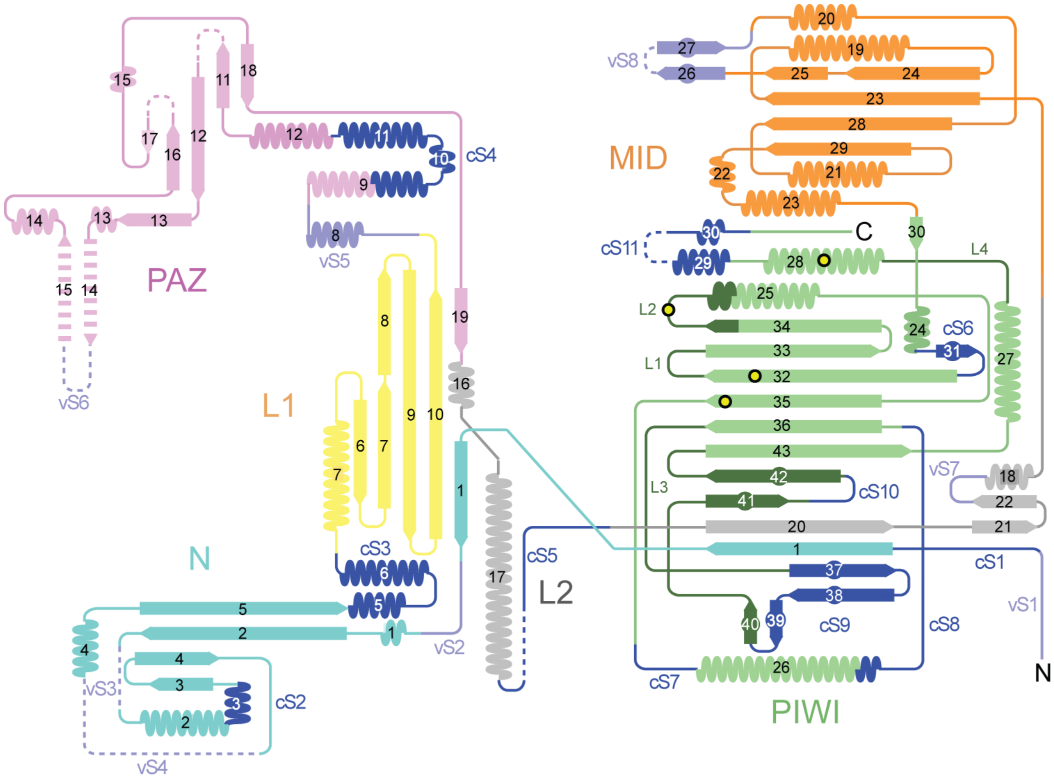
	Δ*xrn1*		Δ*xrn1* Δ*ago1*		Δ*xrn1* sampled	
Total tags	2,513,306		1,739,201			
20–21-nt tags	2,467,140	(98.2)	1,709,290	(98.3)		
Tags mapping to genome	1,796,442	(72.8)	1,322,341	(77.4)		
Tags mapping to ORF sense	589,721	(32.8)	293,400	(22.2)		
Tags mapping to ORF antisense	33,605	(1.9)	19,547	(1.5)		
Tags mapping to Y′ consensus	23,142	(1.3)	2,010	(0.2)	11,239	
Tags mapping to Y′ mRNA	17,972	(77.7)	1,559	(77.6)	8,699	(77.4)
Tags mapping to Y′ antisense	5,170	(22.3)	451	(22.4)	2,540	(22.6)



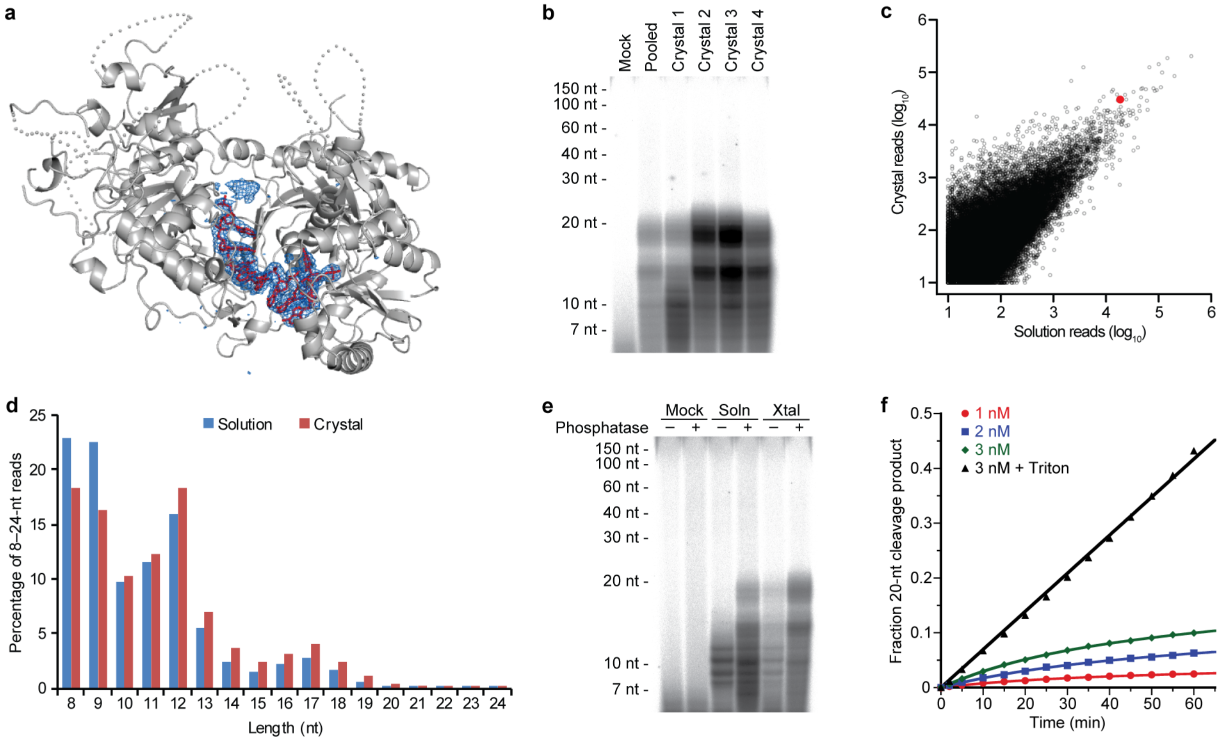
Supplementary Figure 2. Analysis of *S. castellii* degradome libraries
(a) Distribution of degradome tags sequenced from *XRN1*-deletion (Δ*xrn1*) and *XRN1*/*AGO1*-deletion (Δ*xrn1* Δ*ago1*) strains. 20–21-nucleotide sequences were first mapped to the *S. castellii* genome, allowing no mismatches and recovering all hits. The number of genome-mapping reads corresponding to annotated ORFs or the consensus Y′ element are indicated, with percentages shown in parenthesis relative to the parent category. A subset of the complete Δ*xrn1* degradome library was generated to control for library size (Δ*xrn1* sampled; 49% of complete library).
(b) Regression used for library normalization. For each annotated ORF, the number of tags mapping in the sense orientation to annotated exons was calculated by summing the hit-normalized reads. Genes for which either of the strains had zero reads, and ORFs for which there was no annotated homolog in *S. cerevisiae*, were excluded. Red line, result of linear regression performed on log-transformed tag counts between the two strains.
(c) Degradome-cleavage tags mapping to the *S. castellii* Y′ element. The number of tags mapping to the consensus Y′ element are plotted for each position. Above axis, tags from Y′ mRNA; below axis, tags from antisense transcript. For comparison, mapping of Ago1-associated siRNAs, redrawn from Drinnenberg et al. (2009), is shown.
(d) Pairing between cleavage tags and endogenous siRNAs. Degradome tags and siRNAs mapping to the Y′ element were evaluated for antisense overlap between the 5′ ends of the tags and the bodies of the siRNAs. Plotted is the fraction of tags with 5′ ends mapping across from each position within the siRNA. Dashed line, background uniform distribution; gray shading, pairing register expected for AGO cleavage.



Supplementary Figure 3. Representative electron density for KpAGO
Simulated-annealing omit map contoured at 4.2σ around Glu 1013 (left); Lys 939 and Arg 1183 (middle); or Asn 935, Lys 939 and Arg 1183 (right). Protein and RNA segments are drawn and coloured as in Figure 3a.

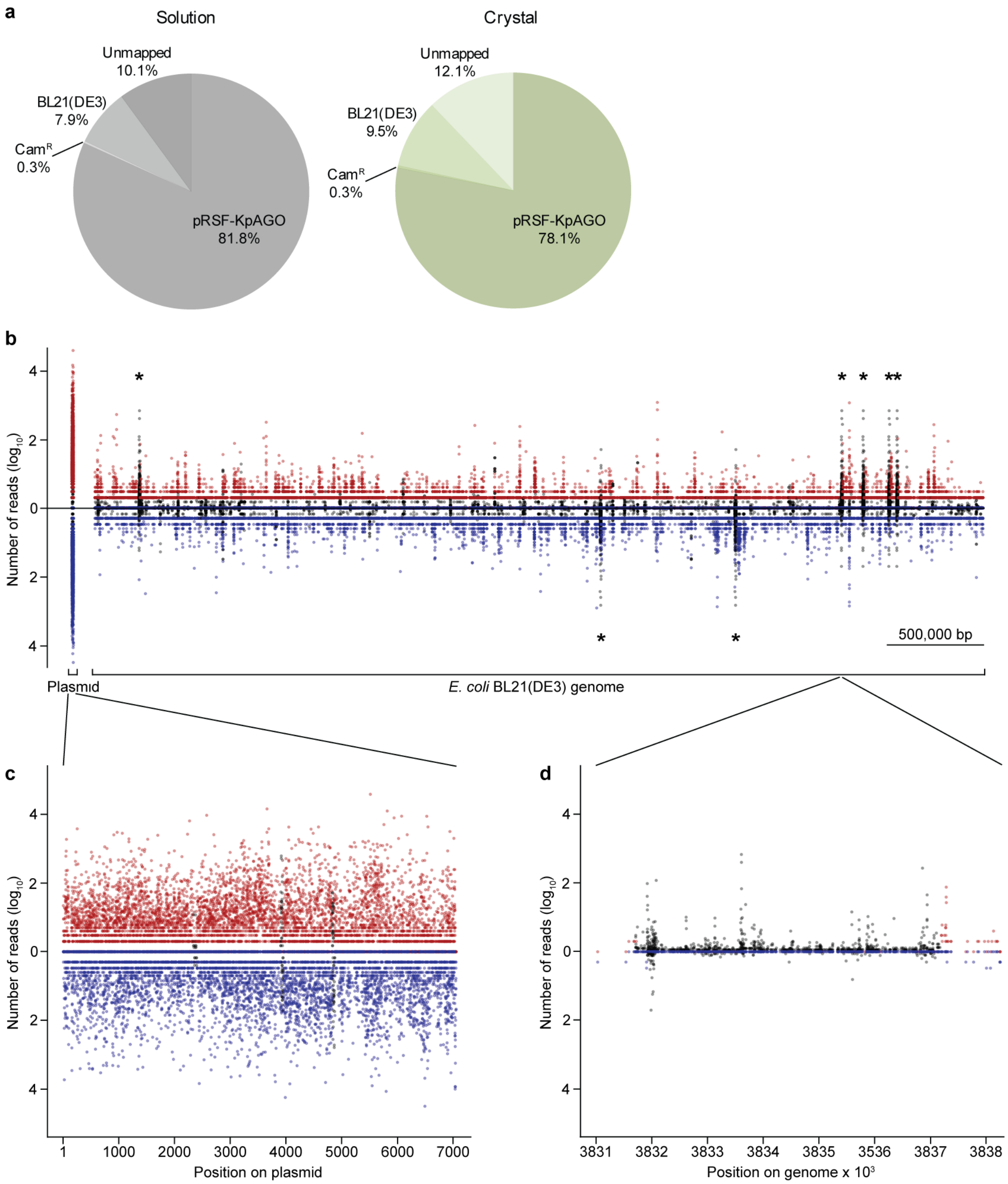


Supplementary Figure 4. Secondary structure and topology diagram of KpAGO
The domain designations and colours are as in Figure 2a. The helices and sheets are shown as wavy lines and arrows, respectively. Loops L1, L2, L3 and L4 are highlighted in dark green; active-site residues are indicated as yellow-filled circles.



Supplementary Figure 5. Analyses of copurifying RNAs

(a) Simulated-annealing omit map (blue) contoured at 3.5σ around the final RNA model (red).
(b) Copurifying RNA associated with a pool of crystals (pooled) and individual crystals (1–4). Copurifying RNA was extracted and analysed as in Figure 2d. The variability between individual crystals implies that RNA associated with the rare crystals that diffracted well might differ from that analysed in other crystals, which might explain why 12-nucleotide passenger-strand cleavage fragments (Supplementary Figure 7) were not observed in the KpAGO structure.
(c) Diverse RNAs copurify with KpAGO (Supplementary Table 2). Shown are the number of reads for a given sequence in small-RNA libraries prepared from soluble (x axis) and crystalline (y axis) KpAGO. Red point, 17-nucleotide copurifying RNA species used in cleavage assays.
(d) Bimodal length distribution of copurifying RNAs. Shown is the percent of reads with the indicated length in small-RNA libraries prepared from soluble and crystalline KpAGO.
(e) Phosphatase-dependence of 5′ end-labelling for most copurifying RNAs. Extracted polynucleotides (or a water-only control, mock) were untreated (–) or treated with phosphatase (+) before 5′ end-labelling with kinase. Otherwise as in panel b.
(f) Kinetics of cleavage guided by a 17-nucleotide copurifying RNA. Reactions containing 1 nM cap-labelled RNA and the indicated concentration of KpAGO were incubated for the indicated time in 1× reaction buffer as in Figure 2f. Reactions were without Triton, unless otherwise indicated. Plotted are average values of two replicates. Solid lines, least-squares fits to a line (Triton condition) or the burst equation (all others).

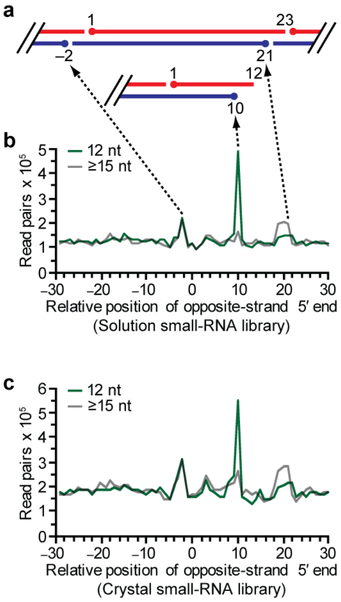


Supplementary Figure 6. Origins of copurifying RNAs

(a) Fraction of reads in small-RNA libraries that map to the KpAGO expression plasmid (pRSF-KpAGO), the chloramphenicol-resistance gene present on pRARE2 in Rosetta2 strains (Cam^R), and the *E. coli* BL21(DE3) genome (Supplementary Table 2).

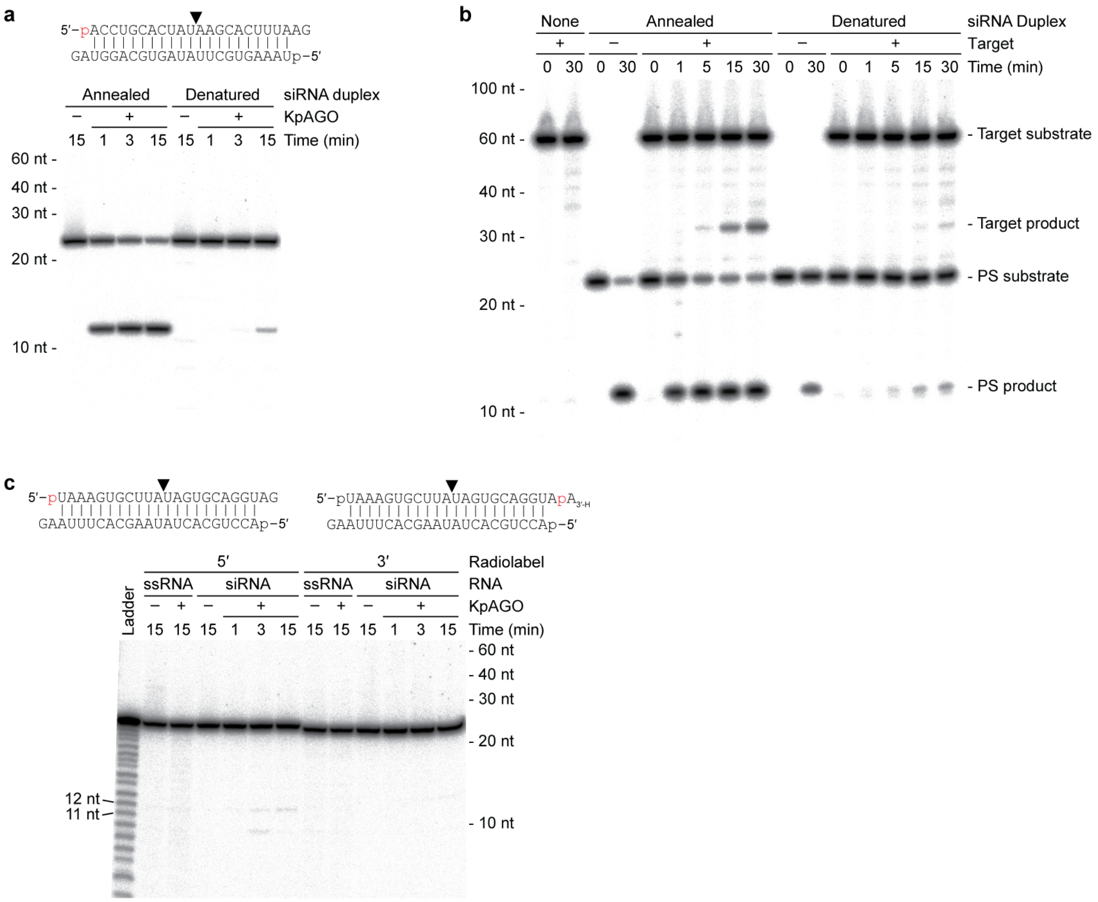
(b) Loci corresponding to copurifying RNAs. Plotted are the number of small-RNA 5′ ends (solution library) mapping to the indicated position on the KpAGO plasmid (left) or *E. coli* genome (right). Red, uniquely aligned plus-strand reads; blue, uniquely aligned minus-strand reads; black, non-uniquely mapping reads (which were hit-normalized, and thus evenly distributed among their corresponding loci). The seven rRNA loci that are abundant sources of copurifying RNAs are indicated (*).

(c) Close-up view of plasmid-mapping reads, plotted as in panel b. The abundance of plasmid-matching copurifying RNAs could be attributed to the high copy number of the plasmid or its propensity to generate double-stranded RNA.
(d) Close-up view of one of seven ribosomal RNA loci, plotted on the same scale as panel c.



Supplementary Figure 7. Relationships between copurifying RNAs from opposite strands
(a) Schematic of relationships expected for 23-nucleotide RNase III cleavage products (top) and those expected for guide RNA–passenger strand cleavage product (bottom).
(b, c) Observed relationships between opposite-strand RNAs extracted from soluble (b) or crystalline (c) KpAGO. Loci of sequences mapping to the plasmid were compared, recording the relative positions of 5′ termini. Plotted in green are the numbers of pairs involving 12-nucleotide RNAs from one strand and ≥15-nucleotide RNAs from the other. Plotted in black are the numbers of pairs involving ≥15-nucleotide RNAs from opposite strands.

The observation that copurifying RNAs matched both strands of the KpAGO expression plasmid, together with the 5′-monophosphate chemistry of the RNA, suggested origins from double-stranded RNAs that were processed by *E. coli* RNase III, which would yield some siRNA-like duplexes with two-nucleotide 3′ overhangs (Xiao et al., 2009). Consistent with this model, the 5′ ends of reads originating from opposite strands tended to be related by a two-nucleotide 3′ overhang (peaks at –2 and +21). Because the 3′ fragment of the cleaved passenger strand is 12 nucleotides (regardless of the initial siRNA duplex length), the enrichment for 12-nucleotide copurifying RNAs (Supplementary Figure 5d) suggested that they might reflect the 3′ fragments of cleaved passenger strands that remained within KpAGO during purification (and then presumably dissociated during formation of well-diffracting crystals). A striking 10-nucleotide phasing observed between 12-nucleotide RNAs and 15–24-nucleotide opposite-strand pairs strongly supported this biogenesis mechanism (green lines, peak at +10).

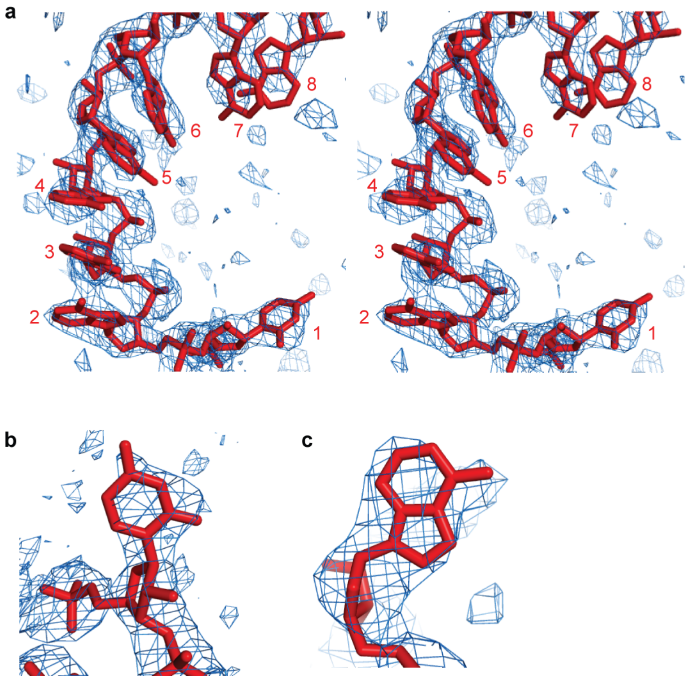


Supplementary Figure 8. Autonomous activities of recombinant KpAGO

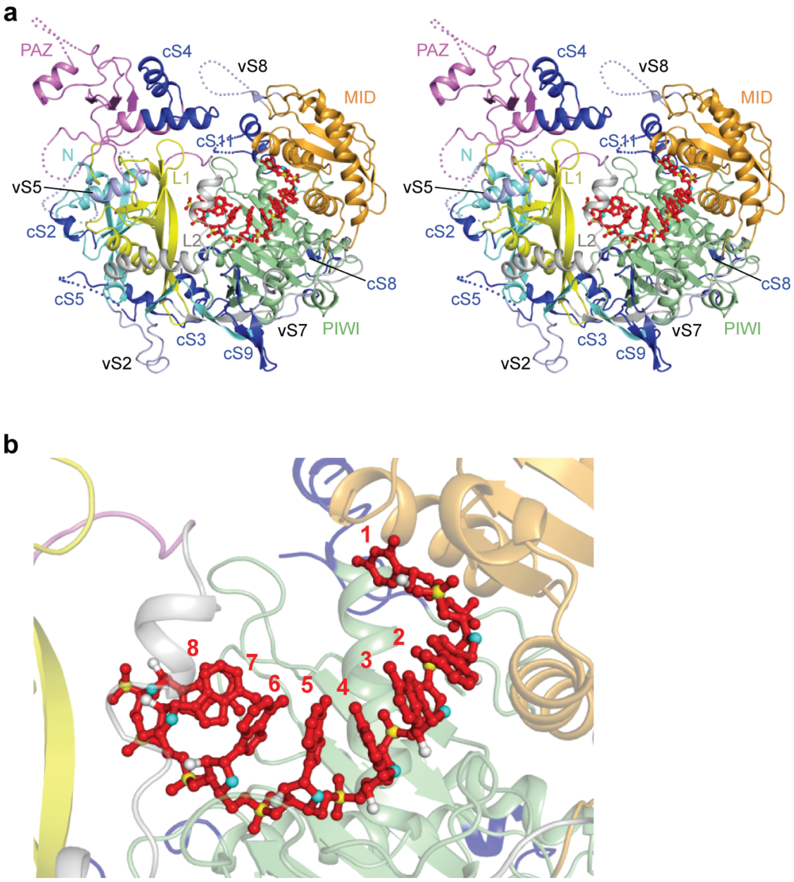
(a) Duplex dependence of passenger-strand cleavage. To confirm that the observed cleavage of labelled passenger strands occurred in the context of an siRNA duplex (rather than denatured single-stranded guide RNA that loaded into KpAGO and cleaved a single-stranded passenger RNA), siRNA duplex was denatured at 90 °C for five min and kept on ice before adding KpAGO. The dramatically reduced passenger-strand cleavage activity indicates that the preferred substrate is an siRNA duplex (with residual cleavage of the denatured duplex attributed to re-annealing during the reaction).

(b) Sequential passenger-strand cleavage and target cleavage. Reactions were as in Figure 2h except without pre-incubation and using end-labelled passenger RNA. The substrates and specific products of passenger-strand (PS) and target cleavage are indicated. The observed kinetics were consistent with rapid passenger-strand cleavage followed by slow release of cleavage fragments to form active RISC, which then cleaves target RNA.

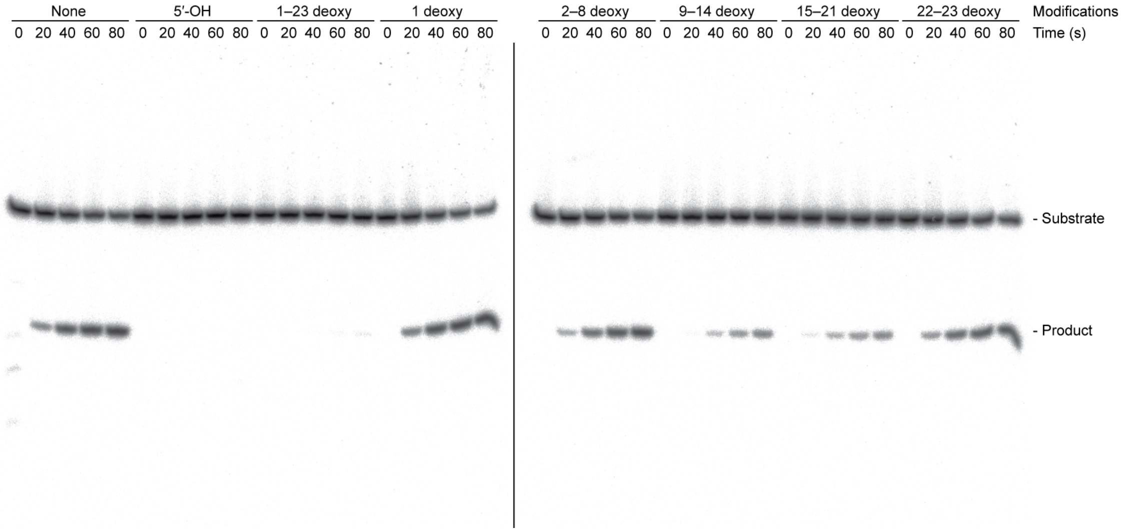
(c) Asymmetric siRNA loading and passenger-strand cleavage by recombinant KpAGO. This experiment was performed together with that shown in Figure 2g and was the same as Figure 2g, except the other strand was labelled. The reduced cleavage of the strand beginning with uridine compared to that of the strand beginning with adenosine (Figure 2g) can be attributed to asymmetric loading of the duplex, with a preference for selecting as the guide the strand beginning with uridine. Denaturing polyacrylamide gel analysis included a lane with partial alkaline-cleavage products (ladder); 11- and 12-nucleotide hydrolysis products are indicated. Note that alkaline-cleavage products (which carry a 2′,3′-cyclic phosphate) migrate approximately one nucleotide faster than do AGO-cleavage products. Passenger-strand cleavage reactions saturated at duplex concentrations far below that of the protein, presumably because most of the protein was already loaded with guide RNA.



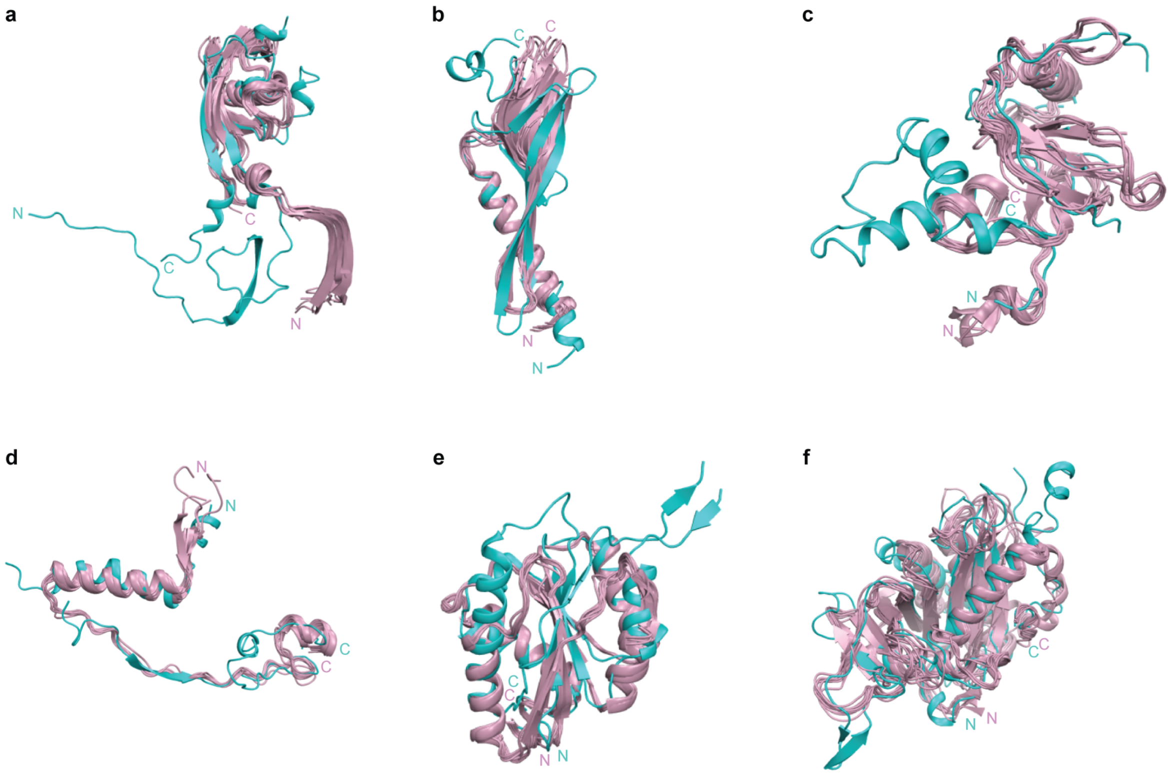
Supplementary Figure 9. Electron density observed for the guide RNA
(a) Stereoview of $F_o - F_c$ electron density map (blue) at 2.6σ around the nucleic-acid-binding channel. The RNA model (red) is drawn as a stick representation with nucleotides numbered from the 5′ end.
(b, c) Close-up view of electron density corresponding to the first (b) and second (c) nucleotides of the guide RNA.



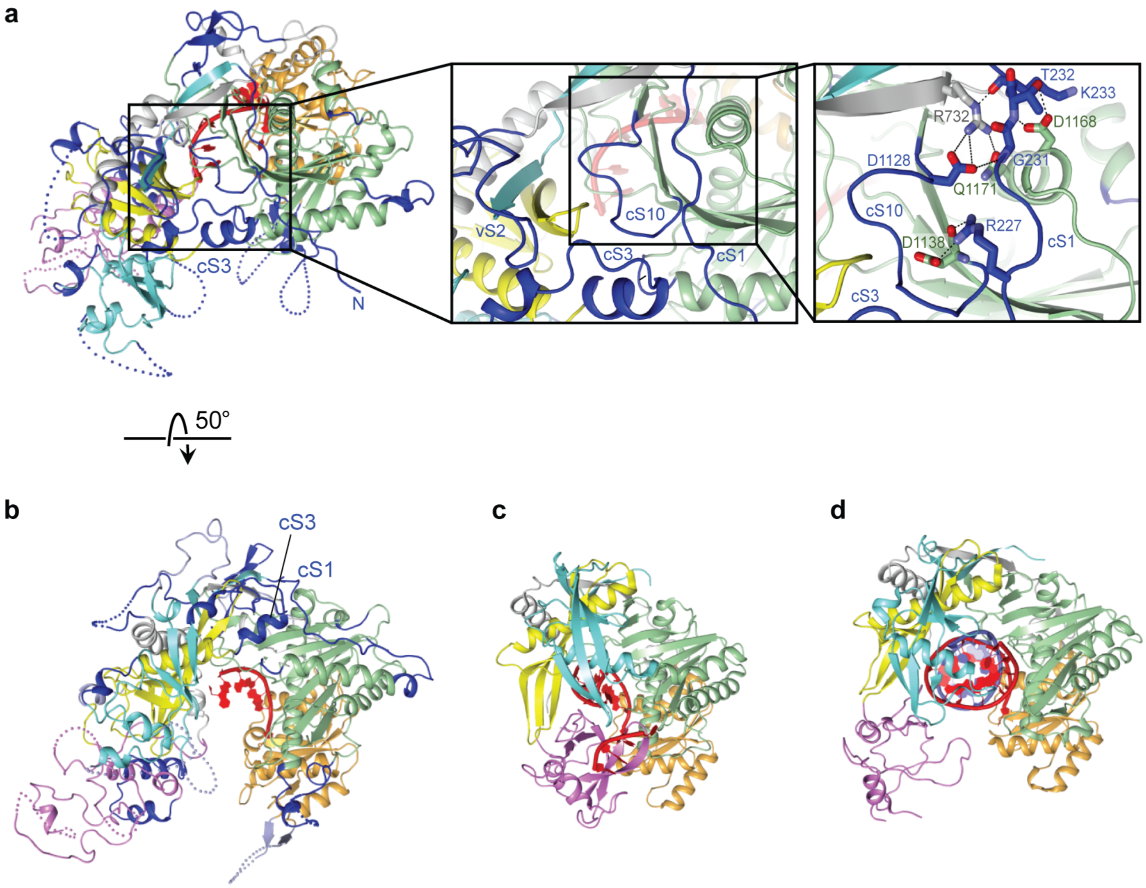
Supplementary Figure 10. Structure of KpAGO with guide RNA
(a) Stereoview of the KpAGO–guide RNA binary complex structure. KpAGO is drawn as in Figure 2a. Guide RNA (red) is drawn as a stick representation with the 2′-OH groups, O4′ atoms and phosphate atoms coloured in white, cyan and yellow, respectively.
(b) Close-up view of the guide RNA (red). Colours are as in panel a, and nucleotides are numbered from the 5′ end.



Supplementary Figure 11. Representative gels of passenger-strand cleavage kinetics. Passenger-strand cleavage assays were performed as described in Figure 3e. Shown is a phosphorimager scan of a denaturing polyacrylamide gel resolving the substrate and cleavage product.



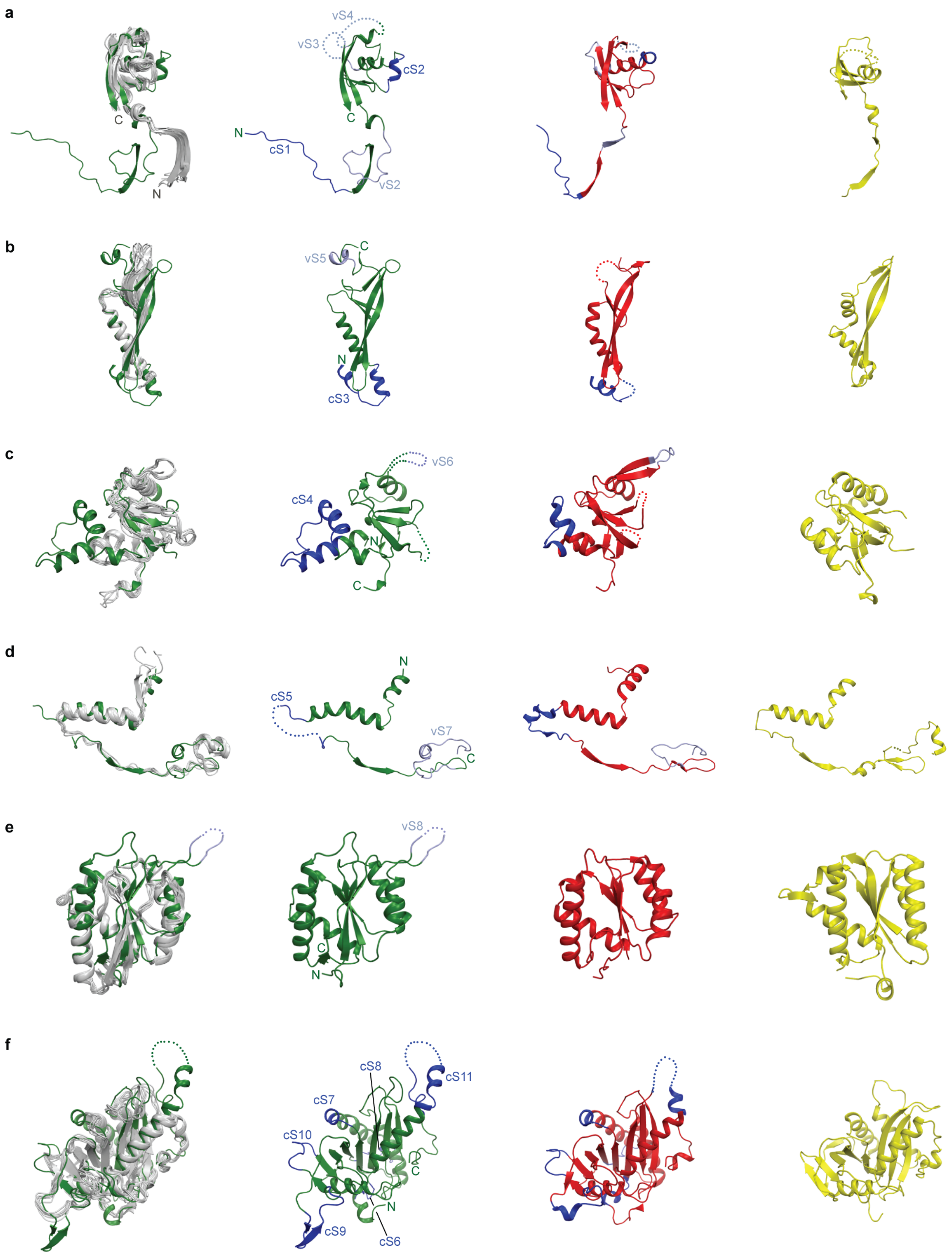
Supplementary Figure 12. Structural alignment of individual domains
Superpositions of the KpAGO (cyan) and TtAGO structures (pink) on their N (a), L1 (b), PAZ (c), L2 (d), MID (e) and PIWI (f) domains. The PDB codes of TtAGO are 3DLB, 3DLH, 3F73, 3HO1, 3HXM, 3HJF, 3HK2, 3HM9 and 3HVR.

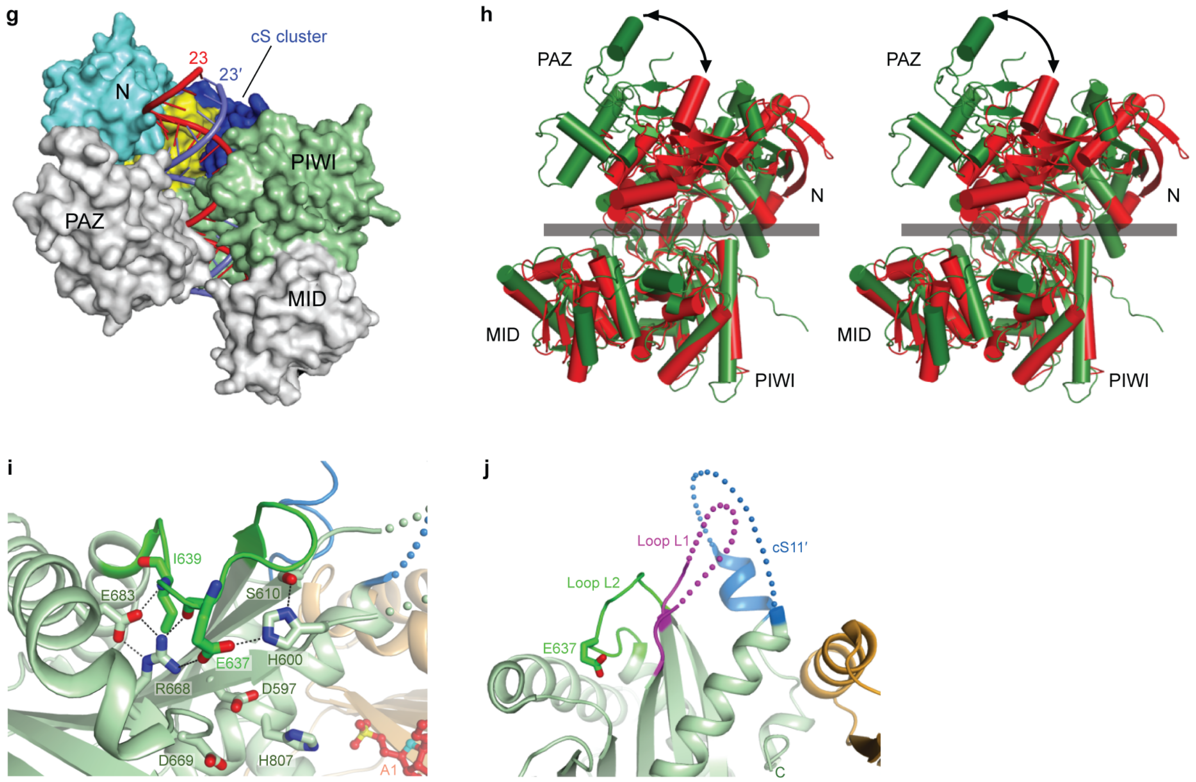


Supplementary Figure 13. The nucleic-acid-binding channels in eukaryotic and prokaryotic AGOs

(a) Cluster of eukaryote-specific insertion segments between the L1 and PIWI domains in the binary KpAGO complex. Arg 227, Arg 732, Asp 1128, Asp 1138, Asp 1168 and Gln 1171 are conserved among eukaryotic AGOs.

(b–d) Comparison of the nucleic-acid-binding channels in the binary KpAGO complex (b), the binary TtAGO complex (c) and the ternary TtAGO complex (d). This view down the nucleic-acid-binding channel shows that there is no obstruction by the KpAGO N domain (cyan), whereas the TtAGO N domain blocks one end of the nucleic-acid-binding channel in both the binary and ternary complexes. Colours are as in Figure 2a. Guide and target strands are coloured in red and slate, respectively.


a
vS4
vS3
cS2
C
C
N
cS1
N
vS2
b
vS5
C
N
cS3
c
vS6
cS4
N
C
d
N
cS5
vS7
C
e
vS8
C
N
f
cS8
cS11
cS7
cS10
C
N
cS9
cS6



Supplementary Figure 14. Structural comparisons with HsAGO2
(a–f) Superpositions of the TtAGO (gray), KpAGO (green), HsAGO2 (red) and *P. furiosus* AGO (yellow) structures on their N (a), L1 (b), PAZ (c), L2 (d), MID (e) and PIWI (f) domains. Constant (cS) and variable (vS) insertion segments in eukaryotic AGOs are coloured in blue and slate, respectively. Disordered regions are drawn as dotted lines. All 11 of the conserved insertion segments in KpAGO, as well as many of the variable insertion segments, are also found in HsAGO2.

(g) HsAGO2 with modelled A-form duplex, revealing an extended and potentially unobstructed nucleic-acid-binding channel. Protein surfaces are rendered, highlighting the cS cluster that extends the channel (blue).

(h) Superposition of KpAGO (green) and HsAGO2 (red) based on the MID-PIWI lobe. The nucleic-acid-binding channel is highlighted (gray bar). A rotation in the relative orientation of the PAZ domain is indicated (black arrows).

(i) Hydrogen-bond network stabilizing the plugged-in loop L2 with the Glu 637 finger, depicted as in Figure 5b.

(j) Conformation of the loop L1 gate, depicted as in Figure 5c. The HsAGO2 segment corresponding to KpAGO cS11 (cS11′) contained four mutations (S824A/S828D/S831D/S834A) with unknown functional consequences.

Chapter 5

Future Directions

Retention and loss of the RNAi pathway

The search for RNAi in budding yeast was initially motivated by the identification of Argonaute homologs in the genome sequences of *S. castellii*, *K. polysporus*, and *C. albicans*. Since the inception of this project, many additional fungal genomes have been sequenced to high coverage. Most of these genomes resemble the *S. cerevisiae* genome in lacking all RNAi components. However, these sequencing efforts have also revealed additional yeasts that contain Dicer and Argonaute homologs and, thus, are likely to have functional RNAi pathways. This expanded collection of budding-yeast Dicer and Argonaute proteins now enables comparative sequence analysis that was not previously possible with such a limited number of homologs, but which may reveal interesting features of these proteins that were formerly missed.

Some budding-yeast species even contain two Dicer or Argonaute genes, which based on synteny appear to be derived from the yeast whole-genome duplication (Wolfe and Shields, 1997). These duplicated genes may be functionally redundant or, alternatively, may have become specialized since their divergence. The RNAi machinery in *Naumovozyma dairenensis* is particularly interesting in having two genes each for Dicer and Argonaute, which could be specialized into two distinct RNA-silencing pathways. Such division-of-labor is seen, for example, in *Drosophila melanogaster*, in which miRNAs act through *Dcr-1*/*Ago-1* while siRNAs act through *Dcr-2*/*Ago-2* (Lee et al., 2004; Forstemann et al., 2007).

On the other hand, some budding-yeast genomes (e.g., *Saccharomyces bayanus*) have a Dicer homolog but no Argonaute homolog, raising the question of the cellular roles of these

Dicers. One possibility is that these Dicer enzymes have lost Dicer activity, i.e., the generation of small-RNA duplexes from long dsRNA. However, sequence analysis indicates that they have maintained the critical features of budding-yeast Dicer, including an intact active site and a C-terminal dsRBD. An intriguing possibility is that these Dicer homologs may be performing non-RNAi functions in these organisms, as was originally speculated for the function of Argonaute homologs in budding yeasts that had no recognizable Dicer (Axelson-Fisk and Sunnerhagen, 2006). Alternatively, these Dicer enzymes may be vestigial relics of a recently lost RNAi pathway, in which case these genes should erode and eventually be lost over evolutionary time (as presumably already happened to the Argonaute genes in these organisms).

Even for those species that have retained the RNAi machinery, it remains to be seen whether they have functional RNAi pathways. One notable example is *C. albicans*, which in addition to Dicer and Argonaute genes has also been shown to have endogenous small RNAs (Drinnenberg et al., 2009). As an obligate diploid with an incomplete sexual cycle, *C. albicans* is difficult to manipulate with traditional gene-knockout strategies (Bennett and Johnson, 2005). Thus, RNAi would be especially useful in *C. albicans* as a tool for gene knock-downs. However, attempts to use hairpin dsRNA to trigger RNAi in *C. albicans* have thus far been unsuccessful (Staab et al., 2011). Whether this is simply a technical issue (e.g., the instability of hairpin DNA constructs) or the biological result of an inactive RNAi pathway is still an open question.

The budding-yeast RNAi pathway provides a transposon defense system that helps to preserve genome integrity, which should provide an evolutionary advantage to those yeasts that retain the pathway. What then can explain the absence of RNAi from so many budding-yeast species, including *S. cerevisiae*? A potential answer to this question came from determining the phenotypic consequences of restoring RNAi to *S. cerevisiae* (Drinnenberg et al., 2011). This analysis revealed that the RNAi pathway is incompatible with an endemic dsRNA viral system known as killer. Because only cells that contain the system are immune to the toxin that it produces, cells without the virus have an evolutionary disadvantage if neighboring cells have acquired the virus. Thus, loss of RNAi could enable acquisition of killer, and this would provide a net selective advantage over RNAi-containing competitors and thereby give rise to RNAi-deficient species. In the future, it will be important to determine how broadly applicable this model is to other instances of RNAi loss, for example in trypanosomatids (Lye et al., 2010).

Additional components of the budding-yeast RNAi pathway

Dcr1 and Ago1 are currently the only known protein components of the budding-yeast RNAi pathway. A central question that remains to be addressed is whether additional proteins are involved in the pathway *in vivo*. Indeed, in all other characterized RNAi pathways at least one other protein functions in the pathway by, for example, facilitating RISC loading (e.g., TRBP in humans and Arb1/Arb2 in *S. pombe*), RISC activation (e.g., C3PO in *Drosophila* and QIP in *Neurospora*), or release of cleaved target RNA (e.g., La in humans) (Chendrimada et al., 2005; Buker et al., 2007; Maiti et al., 2007; Liu et al., 2009; Liu et al., 2011). In certain cases, cofactors are even essential for a functional RNAi pathway, as evidenced by a requirement for TRBP (in addition to Dicer and Argonaute) when reconstituting the human RNAi pathway in *S. cerevisiae* (Suk et al., 2011). Nevertheless, reconstituting the budding-yeast RNAi pathway in *S. cerevisiae* requires expressing only *DCR1* and *AGO1*, raising the possibility that these components alone could comprise the budding-yeast RNAi machinery.

The budding-yeast RNAi pathway differs from all other RNAi pathways in its use of a non-canonical Dicer enzyme, which might obviate the need for a dedicated RISC-loading cofactor. The C-terminal dsRBD of budding-yeast Dicer, which is dispensable for siRNA generation *in vitro* but required for siRNA accumulation *in vivo* (Weinberg et al., 2011), is an attractive candidate for functionally substituting for such a factor. Alternatively, loading of siRNA duplexes into budding-yeast Argonaute may occur spontaneously without the direct involvement of Dicer. This is consistent with *in vitro* studies using purified *K. polysporus* Ago1, in which siRNA duplexes can be autonomously loaded and processed to yield functional RISC (Nakanishi et al., 2012). Even the removal of passenger-strand cleavage fragments and release of cleaved target RNA can occur *in vitro* in the absence of any additional proteins.

Still, additional cofactors may enhance the efficiency of RNAi *in vivo*, and many approaches can be used to identify such cofactors. For example, a comparative genomics approach can be used to identify proteins that are specifically found in the budding-yeast genomes that contain Dicer and Argonaute homologs; these proteins would be obvious candidates for a role in the RNAi pathway. Such an approach was recently applied to RNAi-positive and RNAi-negative trypanosomatids and successfully identified two additional proteins in the *Trypanosoma brucei* RNAi pathway (Barnes et al., 2012). However, even with the availability of additional budding-yeast genome sequences (Gordon et al., 2011), this approach has thus far failed to identify any promising candidates (i.e., protein-coding genes that are not repetitive elements). This suggests that any RNAi cofactor may not be specific to RNAi-containing yeasts, perhaps reflecting an additional non-RNAi-related function in the cell. Such a cofactor might even function in the RNAi pathway that was reconstituted in *S. cerevisiae* using *S. castellii DCR1* and *AGO1*.

Experimental approaches can also be used to search for additional RNAi components. Reasoning that RNAi cofactors might physically associate with Dicer or Argonaute while functioning in the RNAi pathway, candidates can be identified by analyzing the proteins that copurify with Dcr1 or Ago1. This approach has been remarkably successful in the characterization of fungal RNAi pathways, leading to the identification of both the RDRC and ARC complexes in *S. pombe* and the QIP exonuclease in *Neurospora* (Motamedi et al., 2004; Buker et al., 2007; Maiti et al., 2007). Taking advantage of the ease of genetic manipulation in *S. castellii*, homologous recombination can be readily used to add affinity tags to the endogenously expressed proteins, which overcomes potential problems with overexpressing RNAi components or generating suitable antibodies.

A complementary genetics approach can also be used to identify components of the budding-yeast RNAi pathway. Indeed, some of the earliest mechanistic insights into RNAi came from the isolation of mutants with an impaired dsRNA response, which identified homologs of RdRP and Argonaute in fungi, plants, and animals (Cogoni and Macino, 1999; Tabara et al., 1999; Catalanotto et al., 2000; Dalmay et al., 2000; Fagard et al., 2000; Mourrain et al., 2000; Smardon et al., 2000). A genetic selection strategy could be implemented using either the endogenous RNAi pathway in *S. castellii* or the reconstituted pathway in *S. cerevisiae*. For example, the RNAi pathway could be engineered to silence an auxotrophic marker, thus enabling the genetic selection of prototrophic mutants with impaired silencing.

Because many cofactors have already been identified in other RNAi pathways, a candidate-based approach that focuses on the homologs of these cofactors could also reveal additional components of the budding-yeast pathway. Although homologs of dsRBD-containing proteins (e.g., TRBP and R2D2) are absent from budding yeasts, *S. castellii* has a homolog of the La autoantigen as well as two homologs of the *Neurospora* QIP exonuclease. Whether these factors contribute to budding-yeast RNAi can be examined by comparing the efficiency of RNAi in wild-type and knock-out strains using the GFP-reporter systems that were developed in *S. castellii* and *S. cerevisiae* (Drinnenberg et al., 2009; Nakanishi et al., 2012).

Together, these approaches might identify additional proteins that contribute to budding-yeast RNAi *in vivo*. The novel opportunity to use the tools of budding yeast to shed light on the RNAi pathway presents a particularly exciting future direction. Following the discovery of cofactors, a combination of *in vitro* reconstitution and *in vivo* analyses of each step of the RNAi pathway can be used to dissect the specific roles of such proteins. These will hopefully result in a more complete understanding of how the budding-yeast RNAi pathway functions *in vivo*.

Autonomous loading by Argonaute

The structure of budding-yeast Argonaute represented a significant advance in our understanding of eukaryotic Argonaute. These findings were augmented by structures of human Argonaute-2, which revealed conserved structural features that are likely shared by all eukaryotic Argonautes (Elkayam et al., 2012; Schirle and MacRae, 2012). In addition, comparisons of these structures identified some notable differences, including a human-specific binding pocket for GW-rich Argonaute-binding proteins that are involved in the miRNA pathway (Sasaki and Tomari, 2012).

Perhaps one of the biggest surprises that came out of studies of budding-yeast Argonaute was the ability for this purified protein to autonomously load and process siRNA duplexes to generate active RISC (Nakanishi et al., 2012). This differs from the prevailing view that purified Argonaute proteins can assemble RISC using ssRNA guides but not siRNA duplexes (Liu et al., 2004; Rivas et al., 2005). Explaining this discrepancy is critical to our understanding of Argonaute function. One possibility is that there is an intrinsic difference between budding-yeast

Argonaute and human Argonaute. Consistent with this possibility, *K. polysporus* Ago1 adopts a more open conformation than human Argonaute, which might allow the incorporation of bulky siRNA duplexes into yeast Argonaute without an energy-requiring conformational opening of the protein (Sasaki and Tomari, 2012). Alternatively, metazoan Argonautes might also load and process siRNA duplexes autonomously, but such an activity might have been missed by previous analyses that used target cleavage to monitor formation of active RISC. Indeed, recombinant human Argonaute-2 has been shown to cleave the passenger strand of an siRNA duplex *in vitro* (Wang et al., 2009a). This suggests that the failure of metazoan Argonautes to use siRNA duplexes to form RISC *in vitro* might reflect an inability to release inhibitory passenger-strand cleavage fragments rather than a failure to load and cleave siRNA duplexes. Perhaps under the appropriate buffer conditions, even the passenger-strand removal step might proceed in the absence of additional proteins, as shown for *K. polysporus* Ago1 (Nakanishi et al., 2012).

Previous attempts to understand the mechanism of siRNA loading based on the Argonaute structure may have been confounded by the notion that Argonaute alone was not sufficient for loading. In light of the evidence for autonomous loading by Argonaute, it is worth reconsidering how siRNA duplexes could be recognized and loaded by Argonaute within this new paradigm. The PAZ domain has long been assumed to be involved in siRNA loading through recognition of the 2-nt 3′ overhang at one of the termini (Wang et al., 2009b). In addition, the crystal structures of Argonaute have identified the MID–PIWI interface as the 5′-end binding pocket for the guide RNA. This has led to a model for siRNA loading in which the 3′ end of the guide RNA is recognized by the PAZ domain in the context of a 2-nt 3′ overhang, and the 5′ nucleotide of the guide RNA melts away from the duplex to flip into the 5′-end binding pocket (Kawamata and Tomari, 2010).

Since the seed region of the single-stranded guide RNA adopts an A-form-like conformation in eukaryotic RISC, a guide–target duplex can be modeled onto the guide RNA to understand how such a duplex would be accommodated by Argonaute. Moreover, both an siRNA duplex and guide–target duplex adopt the same A-form conformation. Therefore, Argonaute modeled with a guide–target duplex should approximate how an siRNA duplex would be bound by Argonaute with the 5′-end of the guide RNA anchored in the binding pocket. In these models, the 3′ end of the guide RNA is positioned at the intersection of the N–PAZ and MID–PIWI lobes, which is quite distant from the terminus-binding pocket of the PAZ domain. This structural modeling suggests that the PAZ domain would be unable to engage the 3′ end of the guide RNA during autonomous loading. Consistent with PAZ-independent loading, a recent study found that the human Argonaute-2 PAZ domain is dispensable for loading of perfect siRNA duplexes that undergo passenger-strand cleavage (Gu et al., 2012). Instead, the PAZ domain was found to play a role in unwinding small-RNA duplexes during slicing-independent RISC activation. Given these suggestions of PAZ-independent loading *in vivo*, it will be important to determine whether the 2-nt 3′ overhangs at each end of an siRNA duplex contribute to autonomous loading *in vitro*.

If the PAZ domain is not involved in loading, how is an siRNA duplex recognized by Argonaute during autonomous loading? The primary features that distinguish siRNAs from other small-RNA fragments in the cell are duplex structure, length, and 2-nt 3′ overhangs. Therefore, at least one of these features is likely to be important for loading in order to avoid incorporating non-specific RNAs into RISC. Budding-yeast Argonaute more efficiently loads siRNA duplexes *in vitro* compared to the corresponding ssRNA guides, indicating a preference for duplex RNAs (Nakanishi et al., 2012). This duplex character could be recognized simply by its stable A-form conformation, which would facilitate backbone contacts with Argonaute that are maintained in the seed region of the guide after RISC activation. According to this model, the entropic cost of organizing the seed region of an ssRNA guide would impose selectivity for duplex RNAs. The further selectivity for short duplexes might arise from the excess energy required to unwind long duplexes following passenger-strand cleavage. In this way, Argonaute could selectively load small-RNA duplexes without PAZ-domain recognition of the 2-nt 3′ overhang structure. Still, the most complete answer to this loading question awaits the co-crystal structure of Argonaute with an siRNA duplex.

Concluding remarks

S. cerevisiae has long been considered the model budding yeast, but the extent to which this species faithfully represents all budding yeasts has not been thoroughly examined. Indeed, most biological research relies on a few species, which in many cases were chosen for arbitrary reasons. The discovery and characterization of RNAi in budding yeast, and in particular its novel Dicer enzyme, provide a striking example of the limitations of restricting biological research to just a few species. What other biology might we be missing by focusing our studies on only a handful of model organisms, which represent just a small subset of the biological diversity that exists in nature? In this era of high-throughput sequencing, any organism can now become the subject of biological investigation. The distinction between model organism and organism is steadily blurring. Expanding the diversity of organisms that we study can only expand the diversity of phenomena that we discover.

References

Axelson-Fisk, M., and Sunnerhagen, P. (2006). Comparative Genomics: Using Fungi as Models. In Topics in Current Genetics (Heidelberg Germany, Springer), pp. 1-28.

Barnes, R.L., Shi, H., Kolev, N.G., Tschudi, C., and Ullu, E. (2012). Comparative genomics reveals two novel RNAi factors in Trypanosoma brucei and provides insight into the core machinery. PLoS Pathog *8*, e1002678.

Bennett, R.J., and Johnson, A.D. (2005). Mating in Candida albicans and the search for a sexual cycle. Annu Rev Microbiol *59*, 233-255.

Buker, S.M., Iida, T., Buhler, M., Villen, J., Gygi, S.P., Nakayama, J., and Moazed, D. (2007). Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. Nat Struct Mol Biol *14*, 200-207.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature *404*, 245.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Cogoni, C., and Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, 166-169.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543-553.

Drinnenberg, I.A., Fink, G.R., and Bartel, D.P. (2011). Compatibility with killer explains the rise of RNAi-deficient fungi. Science *333*, 1592.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544-550.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Gordon, J.L., Armisen, D., Proux-Wera, E., OhEigeartaigh, S.S., Byrne, K.P., and Wolfe, K.H. (2011). Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci U S A *108*, 20024-20029.

Gu, S., Jin, L., Huang, Y., Zhang, F., and Kay, M.A. (2012). Slicing-Independent RISC Activation Requires the Argonaute PAZ Domain. Curr Biol *22*, 1536-1542.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem Sci *35*, 368-376.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, Y., Tan, H., Tian, H., Liang, C., Chen, S., and Liu, Q. (2011). Autoantigen La promotes efficient RNAi, antiviral response, and transposon silencing by facilitating multiple-turnover RISC catalysis. Mol Cell *44*, 502-508.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science *325*, 750-753.

Lye, L.F., Owens, K., Shi, H., Murta, S.M., Vieira, A.C., Turco, S.J., Tschudi, C., Ullu, E., and Beverley, S.M. (2010). Retention and loss of RNA interference pathways in trypanosomatid protozoans. PLoS Pathog *6*, e1001161.

Maiti, M., Lee, H.C., and Liu, Y. (2007). QIP, a putative exonuclease, interacts with the Neurospora Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes Dev *21*, 590-600.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., *et al.* (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell *101*, 533-542.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Sasaki, H.M., and Tomari, Y. (2012). The true core of RNA silencing revealed. Nat Struct Mol Biol *19*, 657-660.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.

Staab, J.F., White, T.C., and Marr, K.A. (2011). Hairpin dsRNA does not trigger RNA interference in Candida albicans cells. Yeast *28*, 1-8.

Suk, K., Choi, J., Suzuki, Y., Ozturk, S.B., Mellor, J.C., Wong, K.H., MacKay, J.L., Gregory, R.I., and Roth, F.P. (2011). Reconstitution of human RNA interference in budding yeast. Nucleic Acids Res *39*, e43.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Wang, B., Li, S., Qi, H.H., Chowdhury, D., Shi, Y., and Novina, C.D. (2009a). Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. Nat Struct Mol Biol *16*, 1259-1266.

Wang, H.W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009b). Structural insights into RNA processing by the human RISC-loading complex. Nat Struct Mol Biol *16*, 1148-1153.

Weinberg, D.E., Nakanishi, K., Patel, D.J., and Bartel, D.P. (2011). The inside-out mechanism of Dicers from budding yeasts. Cell *146*, 262-276.

Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708-713.

Appendix

Candida albicans Dicer (CaDcr1) is Required for Efficient Ribosomal and Spliceosomal RNA Maturation

Douglas A. Bernstein[a,*], Valmik K. Vyas[a,*], David E. Weinberg[a,b,c], Ines A. Drinnenberg[a,c], David P. Bartel[a,b,c], Gerald R. Fink[a,b]

[a] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142
[b] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[c] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

* These authors contributed equally to this work.

D.E.W. identified active-site mutations in *CDL1* and suggested that *DCR1* may be bifunctional. D.A.B., V.K.V., and I.A.D. performed the experiments. D.A.B., V.K.V., D.P.B., and G.R.F. designed the study. All authors analyzed the data and wrote the manuscript.

Published as:
Bernstein DA*, Vyas VK*, Weinberg DE, Drinnenberg IA, Grisafi PG, Bartel DP, Fink GR (2012). *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc Natl Acad Sci USA* 109:523–8.
** indicates equal contributions*

Candida albicans Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation

Douglas A. Bernstein[a,1], Valmik K. Vyas[a,1], David E. Weinberg[a,b,c], Ines A. Drinnenberg[a,c], David P. Bartel[a,b,c], and Gerald R. Fink[a,b,2]

[a]Whitehead Institute for Biomedical Research, Cambridge, MA 02142; and [b]Department of Biology and [c]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Gerald R. Fink, November 16, 2011 (sent for review October 24, 2011)

The generation of mature functional RNAs from nascent transcripts requires the precise and coordinated action of numerous RNAs and proteins. One such protein family, the ribonuclease III (RNase III) endonucleases, includes Rnt1, which functions in fungal ribosome and spliceosome biogenesis, and Dicer, which generates the siRNAs of the RNAi pathway. The recent discovery of small RNAs in *Candida albicans* led us to investigate the function of *C. albicans* Dicer (CaDcr1). CaDcr1 is capable of generating siRNAs in vitro and is required for siRNA generation in vivo. In addition, *CaDCR1* complements a *Dicer* knockout in *Saccharomyces castellii*, restoring RNAi-mediated gene repression. Unexpectedly, deletion of the *C. albicans CaDCR1* results in a severe slow-growth phenotype, whereas deletion of another core component of the RNAi pathway (*CaAGO1*) has little effect on growth, suggesting that *CaDCR1* may have an essential function in addition to producing siRNAs. Indeed CaDcr1, the sole functional RNase III enzyme in *C. albicans*, has additional functions: it is required for cleavage of the 3′ external transcribed spacer from unprocessed pre-rRNA and for processing the 3′ tail of snRNA U4. Our results suggest two models whereby the RNase III enzymes of a fungal ancestor, containing both a canonical Dicer and Rnt1, evolved through a series of gene-duplication and gene-loss events to generate the variety of RNase III enzymes found in modern-day budding yeasts.

Argonaute | *CDL1* | bifunctional dicer

RNA processing plays pivotal roles in ribosome biogenesis (1), mRNA maturation (2), tRNA synthesis (3), and siRNA generation (4). Each requires the accurate and efficient maturation of precursor RNAs to generate specialized RNA molecules designed to perform distinct cellular tasks.

Dicer is the key ribonuclease important for the generation of siRNAs and related types of small RNAs, which in most eukaryotic organisms have key functions in viral defense, transposon silencing, and cellular gene repression (5–10). Dicer generates siRNA duplexes from long double-stranded RNA (dsRNA). One strand of the duplex is loaded into an effector complex containing Argonaute, where it base pairs to target RNAs, thereby directing gene repression during the process of RNAi. Both Dicer and Argonaute are absent in *Saccharomyces cerevisiae* (8) but present in some other budding-yeast species, including *Saccharomyces castellii* and *Candida albicans*, a fungal pathogen (11). The *S. castellii* Dicer (ScaDcr1) and Argonaute (ScaAgo1) are required for gene silencing but are not essential for viability (11). *C. albicans* has a Dicer-like activity that can process dsRNA into small RNAs in vitro, and in vivo produces small RNAs that bear the chemical signature of molecules generated by a ribonuclease III (RNase III) cleavage (11). However, *C. albicans* Argonaute (CaAgo1) and Dicer (CaDcr1) are reported to be insufficient to induce RNAi-mediated gene silencing in *C. albicans* (12). Thus, the in vivo roles of CaDcr1 and CaAgo1 remain a mystery.

Although the domain structure of budding-yeast Argonaute proteins is similar to that of Argonaute proteins found in other eukaryotes, the budding-yeast Dicers are quite dissimilar from their canonical counterparts found in other fungi and higher eukaryotes (11). These canonical fungal Dicers contain two RNase III domains as well as helicase and Piwi Argonaute Zwille (PAZ) domains (13). By contrast, ScaDcr1 and CaDcr1 have only a single RNase III domain adjacent to a dsRNA-binding domain (dsRBD) and an additional C-terminal dsRBD (11) (Fig. 1*B*). Despite significant differences in domain structure between the budding-yeast and canonical Dicers, the RNase III signature motif and active-site residues important for dsRNA cleavage are conserved in ScaDcr1 and CaDcr1, suggesting that the RNase III cleavage activity is conserved in ScaDcr1 and CaDcr1 (14–16).

Bacteria and some budding yeasts encode a single protein with this RNase III domain, whereas higher eukaryotes often encode several RNase III enzymes that perform distinct tasks in RNA metabolism. Rnt1, the sole RNase III protein in *S. cerevisiae*, plays crucial roles in ribosomal and spliceosomal RNA maturation (17–20). For example, *S. cerevisiae* Rnt1 (SceRnt1) cleaves the 3′ external transcribed spacer (ETS) hairpin from primary 35S rRNA transcript, thereby permitting downstream rRNA processing events (20). Mutants lacking SceRnt1 function grow slowly because they have a severe defect in processing the nascent 35S rRNA transcript, resulting in decreased ribosome biogenesis (20, 21). *S. castellii* encodes two RNase III proteins, ScaRnt1 and ScaDcr1, each thought to perform specialized functions in RNA processing (11). ScaRnt1 has the same domain structure as SceRnt1 and presumably plays analogous roles in ribosomal and spliceosomal RNA maturation. ScaDcr1 contains a second dsRBD not found in SceRnt1 (Fig. 1*B*), suggesting that this domain functions in Dicer-specific activities as seen with *Kluyveromyces polysporus* Dcr1 (16). The in vivo roles for *C. albicans* RNase III homologs are unknown.

To better understand the roles of Dicer and Argonaute in *C. albicans*, we analyzed the consequences of deleting the genes encoding each of these proteins. *C. albicans* mutants lacking *AGO1* grow normally. We were only able to knock out both endogenous copies of *DCR1* when *DCR1* was expressed *in trans* from an inducible promoter. Analysis of strains defective in *DCR1* show that this protein is unusual: *DCR1* encodes a versatile RNase III enzyme that acts as both a Dicer, catalyzing the production of siRNAs, and as an rRNA and snRNA processing enzyme (cleaving the 3′ ETS from unprocessed pre-rRNA and processing the 3′ tail of snRNA U4) like the *S. cerevisiae* Rnt1 enzyme. The Rnt1-like roles of CaDcr1 are likely responsible for the severe slow growth phenotype of mutant strains lacking *DCR1*. Our data suggest that the RNase III enzymes of fungi have evolved through a number of

---

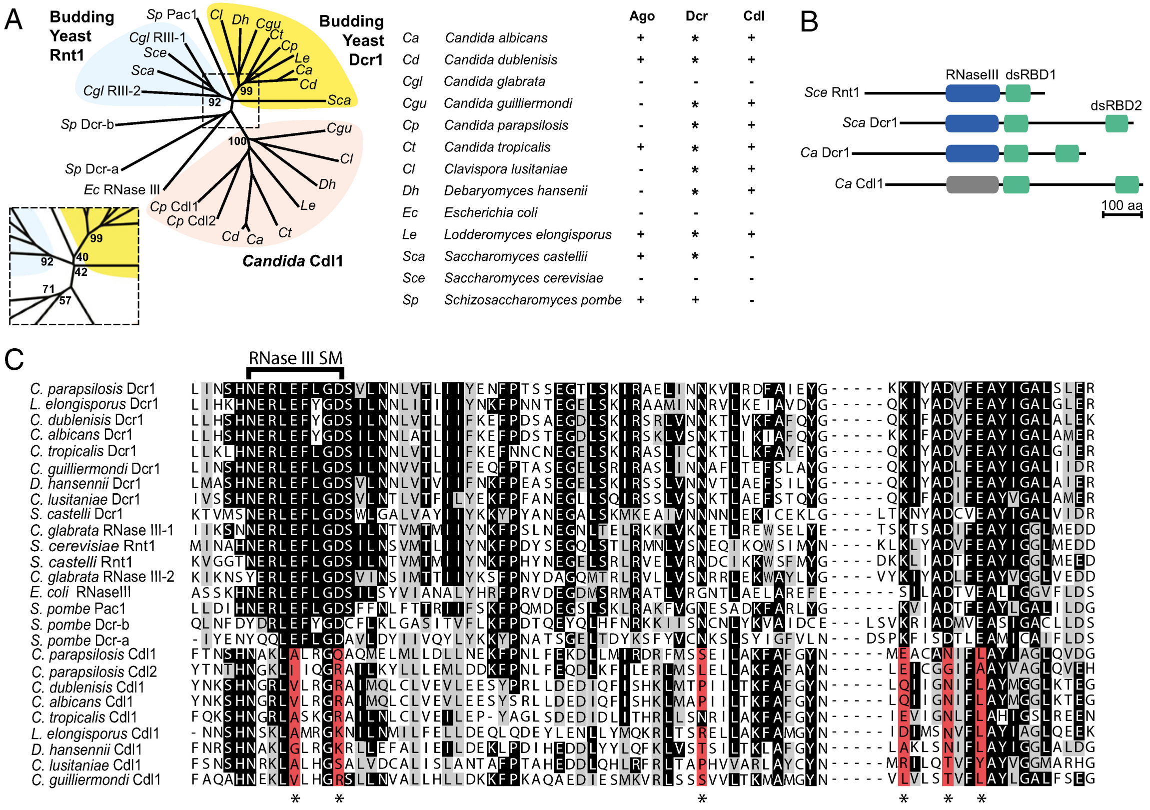
Author contributions: D.A.B., V.K.V., D.P.B., and G.R.F. designed research; D.A.B., V.K.V., and I.A.D. performed research; D.A.B., V.K.V., D.E.W., and I.A.D. contributed new reagents/analytic tools; D.A.B., V.K.V., D.E.W., I.A.D., D.P.B., and G.R.F. analyzed data; and D.A.B., V.K.V., D.E.W., I.A.D., D.P.B., and G.R.F. wrote the paper.

The authors declare no conflict of interest.

[1]D.A.B. and V.K.V. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: gfink@wi.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1118859109/-/DCSupplemental.



		Ago	Dcr	Cdl
Ca	*Candida albicans*	+	*	+
Cd	*Candida dublenisis*	+	*	+
Cgl	*Candida glabrata*	-	-	-
Cgu	*Candida guilliermondi*	-	*	+
Cp	*Candida parapsilosis*	-	*	+
Ct	*Candida tropicalis*	+	*	+
Cl	*Clavispora lusitaniae*	-	*	+
Dh	*Debaryomyces hansenii*	-	*	+
Ec	*Escherichia coli*	-	-	-
Le	*Lodderomyces elongisporus*	-	*	+
Sca	*Saccharomyces castellii*	+	*	-
Sce	*Saccharomyces cerevisiae*	-	-	-
Sp	*Schizosaccharomyces pombe*	+	+	-

Fig. 1. (*A*) Phylogenetic relationship between RNase III enzymes of indicated species. Bootstrap values for key nodes are shown, with central region (boxed) enlarged for clarity. Table indicates the presence (+) or absence (-) of the indicated gene. For Dicer, (*) designates budding-yeast Dicer, whereas + indicates canonical Dicer. (*B*) Domain structure of budding yeast RNase III enzymes. (*C*) Sequence alignment around the RNase III catalytic residues from the protein sequences from *A*. Sequence conservation is labeled in grayscale, with the most highly conserved residues most darkly colored. The six residues (corresponding to SceRnt1 E278, D282, N315, K346, D350, and E353) altered to catalytically inactive residues in the Cdl1 (Candida dicer-like) protein family are labeled (*) and shaded red when inactive. The 9-aa RNase III signature motif is labeled. ClustalW 1.81 with default setting was used to generate both the phylogenetic tree and the sequence alignment.

gene-duplication, neofunctionalization, subfunctionalization, and gene-loss events.

Results

C. albicans AGO1 and DCR1 Mutants. To characterize the in vivo roles of CaDcr1 and CaAgo1, we attempted to construct homozygous deletions of *DCR1* and *AGO1*. The homozygous deletion strains of *AGO1* (*ORF19.2903*) were constructed from wild-type strains by successive transformations with the same deletion construct, with marker removal in between as described in *Materials and Methods*. Homozygous *AGO1* deletion in two different *C. albicans* strain backgrounds had no effect on growth rate (Fig. S1), germ tube formation, or colony morphology.

A search of the *C. albicans* genome for homologs of *S. castellii* Dcr1 revealed two ORFs predicted to encode RNase III domains. One previously identified as the *Candida* Dicer ortholog is encoded by *ORF19.3796* (11), and another is encoded by *ORF19.3773*. Both of these genes are conserved throughout the *Candida* clade and have the same domain structure as ScaDcr1 (Fig. 1 *A* and *B*). Alignment of all of the RNase III domains from the *Candida* clade divided the RNase III-containing proteins into two families, *ORF19.3796* homologs and *ORF19.3773* homologs (Fig. 1 *A* and *C*). Examination of the *ORF19.3773* sequence revealed that amino acids corresponding to E278, D282, N315, K346, D350, and E353 of SceRnt1, which are highly conserved residues involved in metal-ion coordination at the active sites of RNase III enzymes, are altered to residues likely to inactivate RNase activity (Fig. 1*C*) (16). On the basis of these observations, *ORF19.3796* has been named *DCR1*, and *ORF19.3773* has been named *CDL1*, Candida Dicer-like.

Deletions of one copy of *DCR1* in *C. albicans* were constructed; however, deletions of the second copy could not be recovered. Instead, transformations designed to generate a homozygous deletion of *DCR1* resulted in transformants with various genome abnormalities (including aneuploidy and local genome rearrangement events)—all of which retained a functional copy of *DCR1* when examined by Southern blot. On the assumption that complete loss of Dcr1 function might be lethal, we constructed a strain with an ectopic copy of either a maltose-inducible *DCR1* at the *ENO1* locus or a doxycycline (Dox)-inducible *DCR1* at the *ADH1* locus, and one deleted copy of *DCR1*. Under conditions that permitted ectopic *DCR1* expression, we could isolate strains containing disruption of both copies of *DCR1* (Δ*dcr1*/Δ*dcr1*).

The Δ*dcr1*/Δ*dcr1* strains constructed in two different backgrounds had a severe slow-growth phenotype on both solid and liquid media that did not permit transgene expression (Fig. 2 and Fig. S2 *A–N*). The inability to delete both copies of *DCR1* without supplementing the activity from a transgene, coupled with the slow

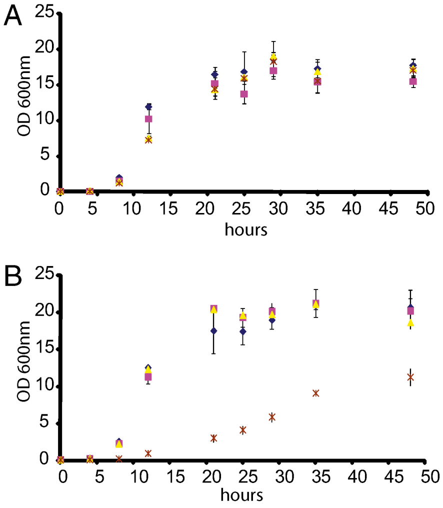


Fig. 2. Effects of *DCR1* on *C. albicans* growth. (*A*) WT [BWP17 (◆)], *dcr1/DCR1* [DAB151 (▲)], *dcr1/DCR1 ENO1/eno1::MAL2p-DCR1* [DAB223 (■)], and *dcr1/dcr1 ENO1/eno1::MAL2p-DCR1* [DAB225 (✕)] grown on a roller drum in synthetic complete (SC) maltose. (*B*) Strains from *A* grown in SC glucose. OD_{600} was used to measure growth. Points plotted are the average ODs of three replicates, and error bars represent 1 SD of the mean.

growth of deletion strains in which the transgene was not induced, suggest that *C. albicans DCR1* encodes an essential function. The slow growth of strains under repressed conditions might be attributable either to the incomplete repression of the ectopic copy of *DCR1* or to residual Dcr1 protein.

C. albicans **Dcr1 Generates Small RNAs in Vivo and in Vitro.** The construction of a strain lacking Dcr1 activity enabled us to determine whether *DCR1* function was required to generate siRNAs. Northern analysis of total RNA prepared from cells grown under repressive conditions revealed that two of the most abundant siRNAs in *C. albicans* (11) were significantly lower in *MAL2p-DCR1, Δdcr1/Δdcr1* than in strains expressing *DCR1* (Fig. 3*A* and Fig. S3 *A* and *B*). Growth of *MAL2p-DCR1, Δdcr1/Δdcr1* in media containing maltose restored siRNA production and actually increased the levels of these siRNAs (Fig. 3*A* and Fig. S3 *A* and *B*) over those found in wild-type cells. Furthermore, a lower abundance siRNA, previously identified by high-throughput sequencing, was only detectable by Northern blotting in *MAL2p-DCR1* containing strains grown on maltose (Fig. S3 *C* and *D*). The increase in siRNA production in the maltose-inducible strains when grown in maltose was likely attributable to increased Dcr1 levels as a result of increased promoter strength, suggesting that Dcr1 is normally limiting for siRNA production. Homozygous deletion of *AGO1* (*Δago1/Δago1*) had no effect on siRNA accumulation as tested by Northern blotting (Fig. S3*E*).

Because *Δdcr1/Δdcr1* mutants had diminished siRNA levels in vivo, we sought to determine whether purified Dcr1 had in vitro dicing activity. Recombinant Dcr1 was purified from *Escherichia coli* and incubated with a 500-bp radiolabeled dsRNA substrate. Purified Dcr1 cleaved the substrate to 22-nt products (Fig. 3*B*), a size matching that of small RNAs sequenced from *C. albicans* (11). These results show that purified CaDcr1 is capable of making siRNAs in vitro and are in agreement with previous work showing that *C. albicans* crude extracts can generate siRNAs from an exogenous dsRNA substrate (11).

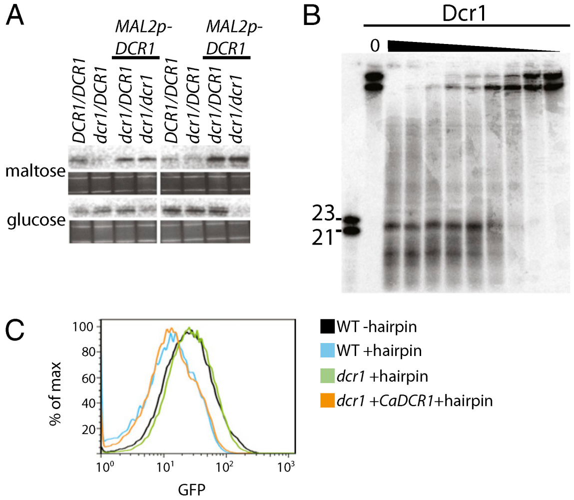


Fig. 3. Functions of *DCR1* in vivo and in vitro. (*A*) Small RNA Northern blot analysis of RNA purified from cells grown overnight in SC and resolved on a denaturing gel. Strains are of indicated genotype, and probe was Northern Primer 1. *Left*: BWP17 background; *Right*: CAI4 background. Below each blot, ribosomal RNA has been visualized with ethidium bromide to show loading. (*B*) In vitro processing of radiolabeled dsRNA by recombinant *C. albicans* Dcr1. Products were resolved on a denaturing gel. Leftmost lane contains size standards, and second lane contains substrate with reaction buffer and no Dcr1. (*C*) *C. albicans DCR1* complements *S. castellii dcr1* for silencing of *GFP*. Strains DPB331 (black), DPB333 (blue), and DPB339 (green) were transformed with pV401, and DPB339 was also transformed with pV406 (orange). Cells were grown in synthetic galactose media and subjected to FACS analysis. Histogram plots represent GFP values for 30,000 cells.

These results raised the question of whether CaDcr1 can act in the RNAi pathway to silence a gene in vivo. Because a gene-silencing system does not yet exist for *C. albicans*, we tested whether CaDcr1 can mediate silencing in *S. castellii*. ScaDcr1 is required to generate the siRNAs that direct gene silencing (11). A recoded *CaDCR1* was transformed into an *S. castellii dcr1* mutant that expresses green fluorescent protein (*GFP*) from the *URA3* promoter and a hairpin against *GFP* from the *GAL1* promoter. The *CaDCR1* complemented the *S. castellii dcr1* mutant, restoring silencing to levels observed in an *S. castellii* strain with a functional *S. castellii DCR1* (Fig. 3*C*). Together, these in vitro and in vivo data demonstrate that *C. albicans DCR1* encodes a gene with Dicer function.

C. albicans DCR1 **Is Required for rRNA Processing.** The requirement of functional CaDcr1 but not CaAgo1 for normal growth suggests that CaDcr1 might have additional roles in *Candida*. In other fungi, mutants lacking either Dicer or Argonaute have minimal effects on growth rates (11, 22, 23). One possibility is that CaDcr1 functions like the SceRnt1 of *S. cerevisiae*, which plays a critical role in the initial steps of ribosome biosynthesis. Decreased expression of SceRnt1 leads to a severe slow-growth phenotype resembling that observed in the *C. albicans Δdcr1/Δdcr1* mutant. Furthermore, the only additional gene in the *Candida* genome that contains an RNase III domain, *CDL1*, is unlikely to have catalytic activity, which leaves CaDcr1 as the only RNase III with the potential to cleave the ribosomal RNA precursor. We performed Northern analysis of *Δdcr1/Δdcr1 MAL2p-DCR1* to compare rRNA processing of cells grown in glucose or permissive media. Strains expressing *DCR1* had no detectable levels of unprocessed rRNA transcripts, whereas the *Δdcr1/Δdcr1* strains grown in restrictive conditions accumulated unprocessed rRNA

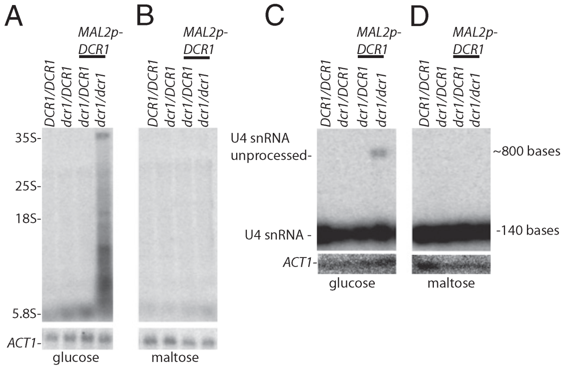


Fig. 4. *DCR1* deletion impact on snRNA and rRNA processing in *C. albicans*. Ten micrograms total RNA prepared from BWP17 background strains grown in indicated carbon source were subjected to Northern analysis probing with pre-rRNA Northern (*A* and *B*), pre-U4 (*C* and *D*), or ACT1 Northern primer (*A*–*D*). Pre-rRNA Northern primer location is depicted in Fig. S5. The pre-rRNA Northern primer hybridizes to the region 21–41 bases past the end of *RDN25*, as depicted in Fig. S5.

transcripts (Fig. 4*A* and Fig. S4 *A*, *C*, and *E*). By contrast, all four strains grown in permissive media had undetectable levels of unprocessed rRNA transcripts (Fig. 4*B* and Fig. S4 *B*, *D*, and *F*). Homozygous deletion of *AGO1* had no effect on rRNA processing (Fig. S4*I*). The rRNA processing defect was unlikely to be an indirect consequence of the loss of an essential gene function, because strains lacking another essential gene (*PES1* under control of *MAL2p*) had severely reduced growth on glucose but did not display the same rRNA processing defect we observed in *Δdcr1*/*Δdcr1 MAL2p-DCR1* (Fig. S4*J*).

Because *S. cerevisiae* Rnt1 plays an additional role in the proper maturation of some snRNA transcripts, we determined whether CaDcr1 also plays an analogous role in snRNA processing. In *S. cerevisiae*, *rnt1* cells accumulate unprocessed U4 transcript (24). On a growth medium containing glucose, we observed an accumulation of unprocessed U4 snRNA transcript in *Δdcr1*/*Δdcr1 MAL2p-DCR1* (Fig. 4*C*) but not in *DCR1*/*DCR1*, *Δdcr1*/*DCR1*, or *Δdcr1*/*DCR1 MAL2p-DCR1* (Fig. 4*D*), indicating that CaDcr participates in U4 snRNA processing in vivo. Correctly processed U4 snRNA also accumulated in nonpermissive conditions, a result similar to that obtained in *S. cerevisiae* (24). This result suggests that *C. albicans* encodes other proteins able to act in a functionally redundant manner, or the small amount of expression in glucose is sufficient for detectable processing.

Discussion

Our findings suggest that CaDcr1, which seems to be the only active RNase III enzyme in *C. albicans*, is versatile, possessing a dicing function that generates siRNAs and additional functions involved in ribosome and spliceosome biogenesis. We attribute the slow-growth phenotype of *Δdcr1*/*Δdcr1* to its ribosome-maturation defects because loss of *CaAGO1* did not affect the growth rate. Knockouts of *S. cerevisiae RNT1* and its *Schizosaccharomyces pombe* homolog *PAC1* also result in severe growth defects that are likely attributable to ribosome-maturation defects similar to those observed in *Δdcr1*/*Δdcr1* (21, 25, 26).

The dual function of the *C. albicans* Dicer is unique among the fungal Dicers that have been analyzed. An *S. castellii* strain lacking Dcr1 function resulted in loss of siRNAs but no obvious loss in viability or discernible phenotype related to growth or morphology (11). The only phenotype that has been noted is the loss of RNAi leads to a compatibility with dsRNA killer viruses (27). Previous work identified small RNAs from *C. albicans* Zorro L1 elements, suggesting a role in genome defense. Because the *Δdcr1*/*Δdcr1* strain has poor viability, the biological roles of siRNAs and other components of the RNAi pathway were best queried using the *Δago1*/*Δago1* strain. Although the siRNAs derived from Zorro L1-like elements imply that these transposable elements are silenced by the RNAi pathway (11), we did not observe any obvious difference in Zorro transposition between *AGO1*/*AGO1* and *Δago1*/*Δago1* strains. However, the Zorro transposition frequency is very low in *Candida*, and slight differences would not have been detected (28).

Our results together with previous work provide insight into the evolution of the budding-yeast RNase III enzymes. Four types of RNase III-like proteins exist in budding yeast (Fig. 1*B*). The *S. cerevisiae* genome encodes a single RNase III enzyme, Rnt1, that is essential for ribosome biosynthesis and also functions in generating spliceosomal RNAs (17, 19, 20), but SceRnt1 does not process dsRNA into siRNAs, consistent with the lack of RNAi in this species (11). *S. castellii* contains two functional RNase III enzymes, one of which likely plays a similar role to SceRnt1 in ribosome and snRNA processing and another, ScaDcr1, which generates siRNAs that guide RNA silencing (11). *K. polysporus* and *Saccharomyces bayanus* both contain syntenic homologs of the *S. castellii DCR1* and *RNT1*. *C. albicans* encodes two proteins with an RNase III domain: Dcr1, which is multifunctional and plays roles in siRNA, snRNA, and rRNA maturation; and Cdl1, which is unlikely to have cleavage activity and whose in vivo role remains a mystery. These two genes are not syntenic with either *DCR1* or *RNT1* from non-*Candida* budding yeasts.

The domain structure of Cdl1 is identical to that of CaDcr1 (Fig. 1*B*), suggesting that it could have originated from a recent gene-duplication event. However, Cdl1 is unlikely to have canonical RNase III activity, because six residues important for RNA cleavage and conserved in other eukaryotic RNase III enzymes are not conserved in Cdl1 homologs. Retention of Cdl1 in the *Candida* clade and conservation of other amino acids in the RNase III domain suggest that it has a function. Catalytically inactive enzyme homologs are common in signaling pathways, and these inactive proteins play roles regulating those pathways (29). Because RNase III proteins act as obligate dimers (14), the potential formation of a heterodimeric complex between CaDcr1 and CaCdl1 is intriguing.

Uncertainties in the branching order within the budding yeast RNase III phylogeny, combined with the absence of informative syntenic relationships, make it difficult to posit a single model that best explains the evolutionary paths of these enzymes. Therefore, we offer two models consistent with our data (Fig. 5). Both assume an ancestral species that had a canonical Dicer (*DICER*) and *RNT1*, as currently observed in the fission yeast *S. pombe*. Model 1 suggests *RNT1* duplication with neofunctionalization to generate a noncanonical Dicer gene (*DCR1*) in a transitional species. The loss of *DICER* left both *RNT1* and *DCR1*, as in present-day *S. castellii*. Loss of *DCR1* and the rest of the RNAi pathway in many budding-yeast lineages (16) left these lineages with only *RNT1*, as observed in present-day *S. cerevisiae* and other members of the *Saccharomyces* complex that lack Argonaute and Dicer homologs. Meanwhile, neofunctionalization of *DCR1* in the *S. castellii*-like ancestor of the *Candida* clade led to the multifunctional enzyme *RNT1*/*DCR1*. *RNT1* loss (and *CDL1* gain through *RNT1*/*DCR1* duplication with neofunctionalization) then generated the genes of *C. albicans*. Model 2 posits early neofunctionalization of an ancestral *RNT1* to generate a transitional species with the multifunctional *RNT1*/*DCR1*. Subsequent *DICER* loss then generated the *Candida*-like budding yeast ancestor, which gained *CDL1* through *RNT1*/*DCR1* duplication with neofunctionalization in the *Candida* lineage. Meanwhile, *DCR1*/*RNT1* duplication with subfunctionalization generated the two RNase III enzymes present in *S. castellii*. This duplication probably did not occur during the whole-genome duplication (WGD) because the synteny typical of paralogs created

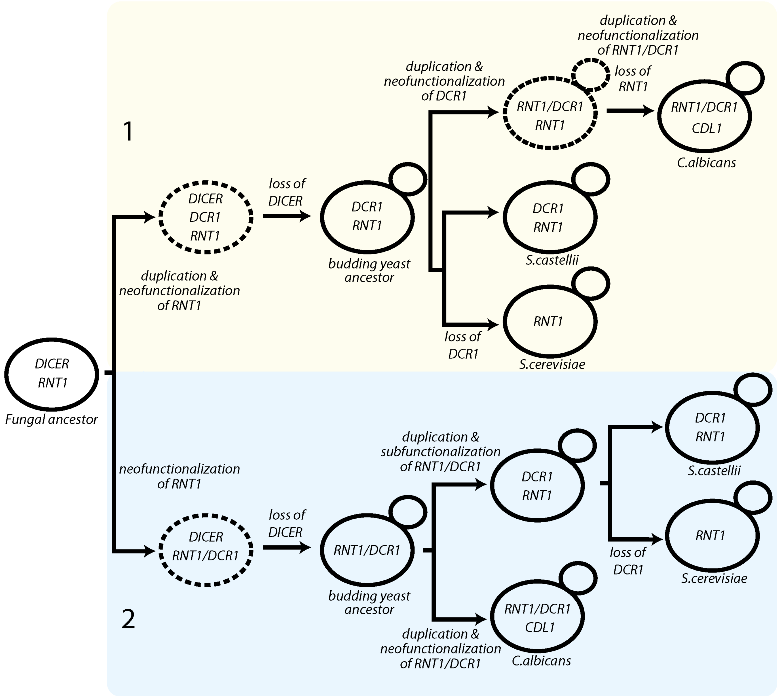


Fig. 5. Two models to explain the evolution of budding yeast DCR1 and RNT1. DICER, canonical Dicer as found in *S. pombe*; DCR1, budding yeast Dicer as found in *S. castellii*, *K. polysporus*, and *S. bayanus*; RNT1, Ribonuclease III as in *S. cerevisiae*; RNT1/DCR1, a multifunctional Dicer found in *C. albicans*; CDL1, Candida Dicer-like from the *Candida* clade.

at the WGD, such as that observed between *Candida glabrata RNT1* paralogs, is not observed for *DCR1* and *RNT1*. The ScaDcr1 RNase III domain is not significantly more similar to the RNase III domains of ScaRnt1 or CaDcr1 (Rnt1/Dcr1), thus we cannot distinguish between these two models with confidence. We favor model 1, because model 2 posits gain and then loss of the second dsRBD during the transitions from *RNT1* to *RNT1/DCR1* and back to *RNT1*. This objection would be mitigated in a hybrid of models 1 and 2, in which

Table 1. Yeast strains

Name	Genotype	Parent	Reference
SC5314	WT clinical isolate	—	35
CAI4	ura3Δ::λimm434 ura3Δ::λimm434	SC5314	36
BWP17	*ura3Δ::λimm434 his1::hisG arg4::hisG* *ura3Δ::λimm434 his1::hisG arg4::hisG*	RM1000	37
VY519	*ago1::FRT/AGO1*	SC5314	This study
VY529	*ago1::FRT/ago1::Nat-flp*	VY519	This study
VY561	*ago1::FRT/ago1::FRT*	VY529	This study
VY523	*ago1::FRT/AGO1*	BWP17	This study
VY537	*ago1::FRT/ago1::FRT*	VY523	This study
VY565	*dcr1::FRT/DCR1*, *adh1::Tetp-DCR1/ADH1*	DAB151	This study
DAB151	*dcr1::FRT/DCR1*	BWP17	This study
DAB157	*dcr1::FRT/DCR1*	CAI4	This study
DAB184	*dcr1::FRT/dcr1-1 adh1::Tetp-DCR1*	DAB151	This study
DAB196	*dcr1::FRT/dcr1-1 adh1::Tetp-DCR1*	DAB157	This study
DAB204	*dcr1::FRT/dcr1-1 adh1::adh1/Tetp-DCR1*	DAB196	This study
DAB232	*dcr1::FRT/dcr1-1 adh1::Tetp-DCR1/ADH1*	DAB184	This study
DAB223	*dcr1::FRT/dcr1-1 eno1::MAL2p-DCR1/ENO1*	DAB151	This study
DAB224	*dcr1::FRT/dcr1-1 eno1::MAL2p-DCR1/ENO1*	DAB157	This study
DAB225	*dcr1::FRT/dcr1-1 eno1::eno1/MAL2p-DCR1*	DAB223	This study
DAB228	*dcr1::FRT/dcr1-1 eno1::eno1/MAL2p-DCR1*	DAB224	This study
JLK713	*pes1::FRT/FRT-MAL2p-PES1*	C665	38
DPB333	*WT + GFP + hairpin*		11
DPB339	*dcr1 + GFP + hairpin*		11

RNT1 is duplicated before *RNT1*/*DCR1* neofunctionalization and then retained in all ancestors of the *Saccharomyces* lineage. Regardless of the model, the evolution of the budding-yeast RNase III enzymes was not a simple linear path but instead required several instances of gene duplication/loss, often with neofunctionalization/subfunctionalization, to arrive at the diversity currently observed.

Materials and Methods

Dicing Assay. Radiolabeled dsRNA (140 pM) was incubated with purified CaDcr1 protein for 30 min at 37 °C, as described previously (11).

Phylogenic Analysis. RNase III domains were identified using the Conserved Domain search of the National Center for Biotechnology Information (30). Alignments of RNase III domains from selected fungi were created with ClustalW on the Biology Workbench [workbench.sdsc.edu (31)], using default parameters. A tree with bootstrap values was created with Clustaltree on Biology Workbench, and Treeview (32). Synteny analysis was performed using the Yeast Gene Order Browser (33) and *Candida* Gene Order Browser (34).

FACS. Cells were grown overnight in appropriate media, diluted in PBS, and subjected to flow cytometry on a BD FACSCalibur with data acquisition using BD CellQuest Pro. Flow cytometry data analysis was done using FlowJo (Treestar).

Strain Construction. All *Candida* strain genotypes in Table 1 were verified by Southern analysis. The KpnI/SacI fragment from plasmid DAB124 was transformed into *C. albicans* strains BWP17 and CAI4 and selected on yeast extract, peptone dextrose media with Nourseothricin (YPD + Nat). Correct integrants were grown in Difco yeast carbon base with bovine serum albumin (YCB + BSA) to induce expression of the flip recombinase and subsequent removal of the *NAT*^{R} gene from the genome generating DAB151 and DAB157. DAB151 and DAB157 were transformed with pV397 integrating *Tetp-DCR1* at the *ADH1* locus. Correct integration in strains DAB151 and DAB157 generated DAB184 and DAB196, respectively. We also transformed DAB151 and DAB157 with pV434 integrating *MAL2p-DCR1* at the *ENO1* locus. Correct integration in strains DAB151 and DAB157 generated DAB223 and DAB224, respectively. A KpnI/SacI fragment of DAB126 was transformed to disrupt the second endogenous copy of *DCR1*, generating *dcr1-1*. The transformation of DAB196 and DAB184 generated DAB204 and DAB232, respectively. An analogous strategy was used to generate strains DAB225 and DAB228 from DAB223 and DAB224, using plasmid DAB177. Although the construction of *DAB225* from *DAB223* resulted in the subsequent loss of the *URA3*, the construction of the Dox-driven expression system resulted in auxotrophically matched strains. Because we see a similar slow-growth phenotype in both constructions, the slow growth of the maltose-inducible strains is not due to the rescue of the uracil auxotrophy. Strains BWP17 and SC5314 were transformed with either pV308 or pV307 (respectively) digested with KpnI/SacII. Correct integrants were grown on YCB + BSA medium to induce expression of FLP, leaving only FRT sites, resulting in strains VY523 and VY519. These strains were subjected to an additional round of transformation and flipping out, resulting in strains VY537 and VY561. *S. castellii* strains DPB333 and DPB339 were transformed using the lithium acetate method, with either pV401 or pV406 for complementation analysis, and selected on YPD + Nat.

Additional methods included in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Kotaro Nakanishi, Julia Köhler, Joachim Morschhäuser, and Paula Sundstrom for plasmids and strains, Julia Köhler for advice, and members of the G.R.F. and D.P.B. laboratories for helpful suggestions. This work was funded by American Cancer Society Grant PF-09-072-01-MBC (to D.A.B.), National Institutes of Health (NIH) National Research Service Award F32 AI729353 (to V.K.V.), a Herman Sokol Fellowship (to V.K.V.), a Boehringer-Ingelheim Fonds fellowship (to I.A.D.), a National Science Foundation graduate fellowship (to D.E.W.), and NIH Grants GM040266 (to G.R.F.) and GM061835 (to D.P.B.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

1. Venema J, Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet* 33:261–311.
2. Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336.
3. Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. *Genes Dev* 24: 1832–1860.
4. Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* 431:343–349.
5. Zamore PD, Haley B (2005) Ribo-gnome: The big world of small RNAs. *Science* 309: 1519–1524.
6. Ding SW, Voinnet O (2007) Antiviral immunity directed by small RNAs. *Cell* 130: 413–426.
7. Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. *Cell* 136: 656–668.
8. Nakayashiki H, Kadotani N, Mayama S (2006) Evolution and diversification of RNA silencing proteins in fungi. *J Mol Evol* 63:127–135.
9. Hall TM (2005) Structure and function of argonaute proteins. *Structure* 13:1403–1408.
10. Czech B, Hannon GJ (2011) Small RNA sorting: Matchmaking for Argonautes. *Nat Rev Genet* 12:19–31.
11. Drinnenberg IA, et al. (2009) RNAi in budding yeast. *Science* 326:544–550.
12. Staab JF, White TC, Marr KA (2011) Hairpin dsRNA does not trigger RNA interference in *Candida albicans* cells. *Yeast* 28:1–8.
13. Nowotny M, Yang W (2009) Structural and functional modules in RNA interference. *Curr Opin Struct Biol* 19:286–293.
14. MacRae IJ, Doudna JA (2007) Ribonuclease revisited: Structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol* 17:138–145.
15. MacRae IJ, Zhou K, Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 14:934–940.
16. Weinberg DE, Nakanishi K, Patel DJ, Bartel DP (2011) The inside-out mechanism of Dicers from budding yeasts. *Cell* 146:262–276.
17. Abou Elela S, Ares M, Jr. (1998) Depletion of yeast RNase III blocks correct U2 3′ end formation and results in polyadenylated but functional U2 snRNA. *EMBO J* 17: 3738–3746.
18. Lamontagne B, Larose S, Boulanger J, Elela SA (2001) The RNase III family: A conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Curr Issues Mol Biol* 3:71–78.
19. Chanfreau G, Elela SA, Ares M, Jr., Guthrie C (1997) Alternative 3′-end processing of U5 snRNA by RNase III. *Genes Dev* 11:2741–2751.
20. Elela SA, Igel H, Ares M, Jr. (1996) RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* 85:115–124.
21. Chanfreau G, Legrain P, Jacquier A (1998) Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol* 284:975–988.
22. Provost P, et al. (2002) Dicer is required for chromosome segregation and gene silencing in fission yeast cells. *Proc Natl Acad Sci USA* 99:16648–16653.
23. Wang X, et al. (2010) Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. *Genes Dev* 24:2566–2582.
24. Allmang C, et al. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* 18:5399–5410.
25. Iino Y, Sugimoto A, Yamamoto M (1991) *S. pombe* pac1+, whose overexpression inhibits sexual development, encodes a ribonuclease III-like RNase. *EMBO J* 10:221–226.
26. Xu HP, Riggs M, Rodgers L, Wigler M (1990) A gene from *S. pombe* with homology to *E. coli* RNAse III blocks conjugation and sporulation when overexpressed in wild type cells. *Nucleic Acids Res* 18:5304.
27. Drinnenberg IA, Fink GR, Bartel DP (2011) Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* 333:1592.
28. Goodwin TJ, Ormandy JE, Poulter RT (2001) L1-like non-LTR retrotransposons in the yeast *Candida albicans*. *Curr Genet* 39:83–91.
29. Pils B, Schultz J (2004) Inactive enzyme-homologues find new function in regulatory processes. *J Mol Biol* 340:399–404.
30. Marchler-Bauer A, et al. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
31. Subramaniam S (1998) The Biology Workbench—a seamless database and analysis environment for the biologist. *Proteins* 32:1–2.
32. Page RD (1996) TreeView: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357–358.
33. Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15: 1456–1461.
34. Fitzpatrick DA, O'Gaora P, Byrne KP, Butler G (2010) Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. *BMC Genomics* 11:290.
35. Gillum AM, Tsay EY, Kirsch DR (1984) Isolation of the *Candida albicans* gene for orotidine-5′-phosphate decarboxylase by complementation of *S. cerevisiae* ura3 and *E. coli* pyrF mutations. *Mol Gen Genet* 198:179–182.
36. Fonzi WA, Irwin MY (1993) Isogenic strain construction and gene mapping in *Candida albicans*. *Genetics* 134:717–728.
37. Wilson RB, Davis D, Mitchell AP (1999) Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J Bacteriol* 181:1868–1874.
38. Shen J, Cowen LE, Griffin AM, Chan L, Köhler JR (2008) The *Candida albicans* pescadillo homolog is required for normal hypha-to-yeast morphogenesis and yeast proliferation. *Proc Natl Acad Sci USA* 105:20918–20923.

Supporting Information

Bernstein et al. 10.1073/pnas.1118859109

SI Materials and Methods

Northern Blots. Log phase cultures grown in appropriate media at 30 °C were harvested, and total RNA was isolated using acid phenol purification followed by DNase treatment using Ambion's Turbo DNA-free kit. Between 5 and 10 µg total RNA was loaded onto Mops buffered agarose (Ambion) denaturing gel and run for 2 h at 80 V. RNA was transferred to Hybond N+ membrane and UV cross-linked with 12 J/cm^2. Blots were washed with 10 mL Sigma Perfect Hyb Plus. Oligonucleotide probes complementary to the region of interest were end-labeled with adenosine 5′-triphosphate-^{32}P by T4 polynucleotide kinase. Blots were incubated in 10 mL Sigma Perfect Hyb Plus with radiolabeled probe overnight at 57 °C, washed, and visualized by phosphorimaging. Oligonucleotide sequences of probes are below.

sRNA Northern Blots. RNA was harvested as above. Twenty micrograms of total RNA was subjected to Northern analysis as described previously (1).

Dcr1 Protein Purification. A recoded *CaDCR1* gene was cloned into a SUMO-RSF-Duet vector creating an *N*-[His-Sumo-protease site-Dcr1] construct. *Escherichia coli* strain BL21(DE3)pLysS was transformed with this vector, and cultures were grown to an OD of 0.5 at 37 °C. Dcr1 expression was induced for 4 h at 37 °C with 1 mM isopropyl β-D-thiogalactopyranoside, and cells were collected and frozen in lysis buffer [10 mM Tris (pH 8.0), 300 mM NaCl, 1 mM β-mercaptoethanol, 20 mM imidazole, and 10% glycerol]. Cells were lysed by sonication, and lysates were cleared by centrifugation. Supernatant was passed over Ni-NTA agarose beads, and bound proteins were eluted with elution buffer [10 mM Tris (pH 8.0), 300 mM NaCl, 1 mM β-mercaptoethanol, 100 mM imidazole, and 10% glycerol]. Eluted protein was dialyzed into lysis buffer, and the His and Sumo tags were cleaved with the SUMO protease kit (Invitrogen). The cleavage products were passed over Ni-NTA agarose beads, and the flow-through was collected.

Growth Curve: Liquid Growth Curve. Saturated cultures that were grown overnight in permissive media were diluted to a starting OD$_{600}$ of 0.1. Strains were grown in synthetic complete media at 30 °C containing either glucose or maltose, and OD$_{600}$ were measured and plotted. Points plotted are the average of three replicates, and error bars represent 1 SD from the mean of these three replicates.

Solid Media Growth. Saturated overnight cultures were serially diluted onto synthetic complete media containing either glucose or maltose as a carbon source and grown for 2 d at 30 °C.

Plasmids. A recoded *CaDCR1* PCR fragment with XbaI/XhoI sites was cloned into p415-TEF (2) to create pV352. pV406 was created by cloning a Mung-bean-nuclease-blunted NotI fragment from pAG36 (3) containing the *NAT*^{R} into the HpaI site of pV352. pV401 was created by cloning a Mung-bean-nuclease-blunted NotI fragment of pAG36 containing *NAT*^{R} into the HpaI site of p415-TEF. The pV397 doxycycline (Dox)-inducible expression construct for *CaDCR1* in *Candida* was created by replacing the SalI/BglII fragment containing *GFP* from pNIM1 (4) with a SalI/BglII PCR fragment containing *CaDCR1*. Maltose-induced *CaDCR1* was produced from pV434, which contains a HindIII/PstI PCR fragment of *CaDCR1* at the HindIII/PstI sites of pV356. pV356 was created by replacing the HindIII/XhoI flanked *CaENO1* promoter from pENO1-EGFP3 (5) with a HindIII/XhoI fragment containing the *CaMAL2* promoter. The *NAT*^{R}*-FLP* disruption cassette for *CaAGO1*, pV307, was created by cloning 1,100 bp immediately upstream of the ORF into the KpnI/ApaI sites of pJK863 (6) and 2,000 bp downstream into the NotI/SacII. pV308, which contains a *URA3-FLP* disruption cassette, was created by cloning the same fragments at the same sites of pSFU1 (7). DAB124 was created by cloning 2,000 bp upstream and downstream of *DCR1* into JK863 as was for pV307, and DAB126 is DAB124 with the last 500 bp of *DCR1* in the place of the 2,000 bp downstream, and *NAT*^{R} from pJK795; DAB177 is DAB124 with *URA3* in place of *NAT*^{R}.

Probe Sequences.

pre-U4 Northern primer: 5′ CTCGGACGAATCCTCACTGATACACG
sRNA Northern primer 1: 5′ GATGTCCTGGAAATGGATACCA
sRNA Northern primer 2: 5′ TATCCTGATGATTTCAATATCT
sRNA Northern primer 3: 5′ ATTCTTGATCACATCATCAGCT
pre-rRNA Northern primer: 5′ GCCTATGCTAAACAGACCCACC
ACT1 Northern primer: 5′ CCATATCGTCCCAGTTGGAAA

1. Drinnenberg IA, et al. (2009) RNAi in budding yeast. *Science* 326:544–550.
2. Mumberg D, Müller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156:119–122.
3. Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 15:1541–1553.
4. Park YN, Morschhäuser J (2005) Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. *Eukaryot Cell* 4:1328–1342.
5. Staab JF, Bahn YS, Sundstrom P (2003) Integrative, multifunctional plasmids for hypha-specific or constitutive expression of green fluorescent protein in *Candida albicans*. *Microbiology* 149:2977–2986.
6. Shen J, Guo W, Köhler JR (2005) CaNAT1, a heterologous dominant selectable marker for transformation of *Candida albicans* and other pathogenic *Candida* species. *Infect Immun* 73:1239–1242.
7. Morschhäuser J, Michel S, Staib P (1999) Sequential gene disruption in *Candida albicans* by FLP-mediated site-specific recombination. *Mol Microbiol* 32:547–556.

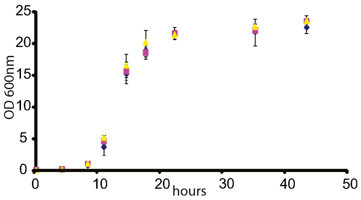


Fig. S1. Deletion of *AGO1* does not impact growth rate. WT [BWP17 (◆)], *ago1/AGO1* [VY523 (▲)], and *ago1/ago1* [VY537 (■)] were grown on a roller drum in synthetic complete (SC) glucose. OD_{600} was used to measure growth. Points plotted are the average ODs of three replicates, and error bars represent 1 SD of the mean.

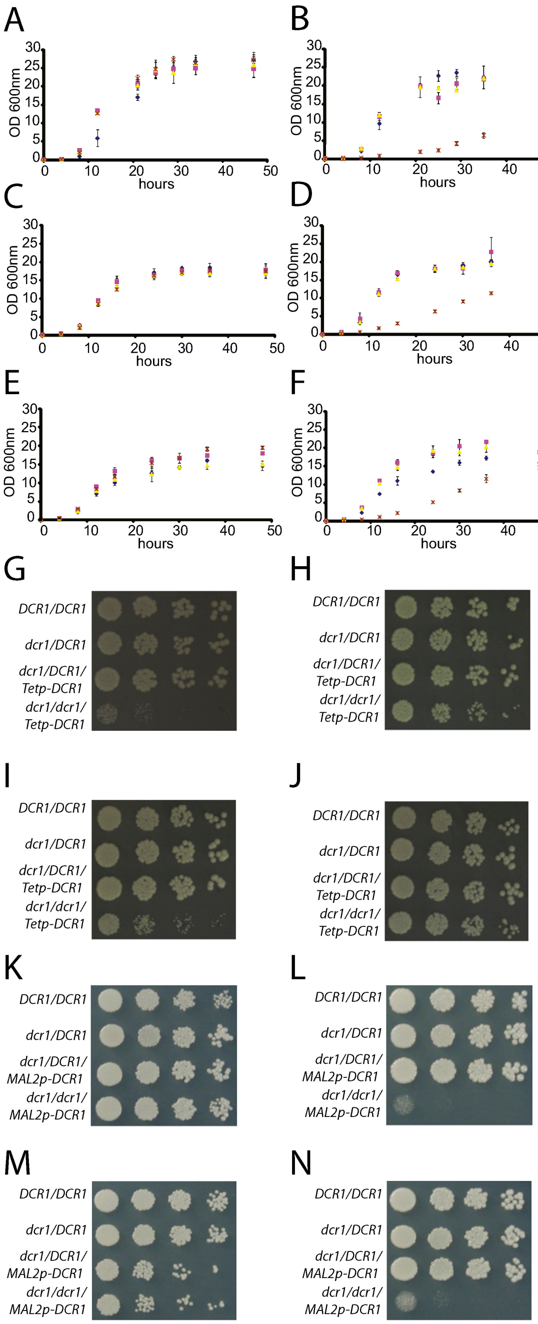


Fig. S2. *DCR1* deletion impact on *Candida albicans* growth. (*A*) CAI4 (◆), DAB157 (▲), DAB224 (■), and DAB228 (✕) grown on a roller drum in synthetic complete (SC) maltose. (*B*) CAI4 (◆), DAB157 (▲), DAB224 (■), and DAB228 (✕) grown on a roller drum in SC glucose. (*C*) BWP17 (◆), DAB151 (▲), DAB184 (■), and DAB232 (✕) grown on a roller drum in SC glucose with Dox. (*D*) BWP17 (◆), DAB151 (▲), DAB184 (■), and DAB232 (✕) grown on a roller drum in SC glucose. (*E*) CAI4 (◆), DAB157 (▲), DAB196 (■), and DAB204 (✕) grown on a roller drum in SC glucose with Dox. (*F*) CAI4 (◆), DAB157 (▲), DAB196 (■), and DAB204 (✕) grown on a roller drum in SC glucose. Points plotted are the average ODs of three replicates, and error bars represent 1 SD of the mean. (*G*) BWP17 strains of indicated genotype on SC glucose. (*H*) Strains from *G* on SC glucose with Dox. (*I*) CAI4 strains of indicated genotype on SC glucose. (*J*) Strains from *I* on SC glucose with Dox. (*K*) BWP17 background strains of indicated genotype on SC maltose. (*L*) Strains as in *K* on SC glucose. (*M*) CAI4 strains of indicated genotype on SC maltose. (*N*) strains as in *M* on SC glucose.

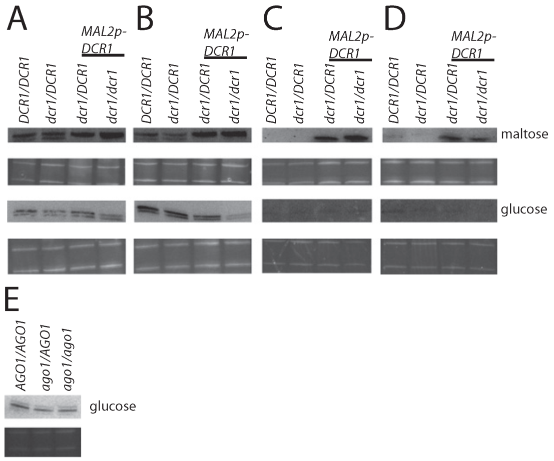


Fig. S3. Small RNA production in vivo by CaDcr1. (*A*) Small RNA Northern blot analysis of RNA purified from cells grown in SC maltose or glucose resolved on a denaturing gel. Strains are of indicated genotype, panel is BWP17 background. Blot probed with sRNA Northern primer 2. (*B*) Northern blot analysis of RNA samples were purified after being grown in SC glucose or maltose and resolved on a denaturing gel. Strains are of indicated genotype, panel is CAI4 background. Blot probed with sRNA Northern primer 2. Below each blot, ribosomal RNA has been visualized with ethidium bromide to show equal loading. (*C* and *D*) Similar to *A* and *B*, but a lower-abundance sRNA was probed with sRNA Northern primer 3. (*E*) Northern blot analysis of sRNA in BWP17 background with sRNA Northern primer 1.

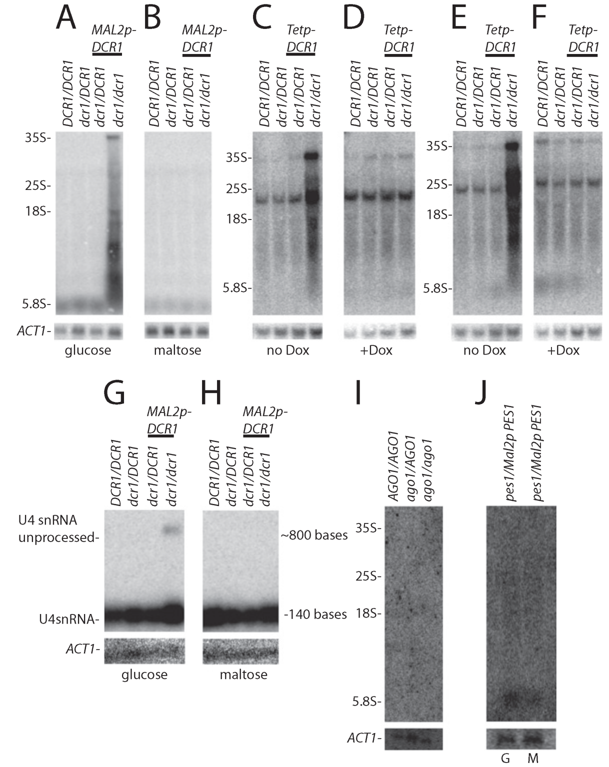


Fig. S4. *DCR1* deletion impact on rRNA processing in *Candida albicans*. (*A–F*) Ten micrograms of total RNA was subjected to Northern analysis, probing for unprocessed rRNA using rRNA Northern probe. Strains are of indicated genotype. *A*, *B*, *E*, and *F* are CAI4 background; *C* and *D* are BWP17. Growth condition is synthetic complete with carbon source noted below. Ribosomal RNA sizes are labeled. Blots were stripped and probed for *ACT1* transcript as a loading control. (*G* and *H*) Northern analysis of U4 processing using pre-U4 Northern probe and ACT1 Northern probe. Strains are of indicated genotype panel in the CAI4 background, grown in the condition indicated below. (*I*) Northern analysis of 10 μg RNA prepared from BWP17 strains of indicated genotype grown in SC glucose probed with pre-rRNA Northern primer or ACT1 northern primer. (*J*) Northern of 7 μg total RNA from JLK713 grown in glucose (G) or maltose (M) with pre-rRNA Northern primer or ACT1 Northern primer.

Fig. S5. Position of pre-rRNA Northern primer (used in Fig. 4 and Fig. S4) with respect to *RDN25* sequence. Numbers indicate distance past the end of *RDN25*.

Curriculum vitae

David E. Weinberg

Education

2007–2013: PhD in Biology, Massachusetts Institute of Technology, Cambridge, MA

2003–2007: BS/MS in Molecular Biophysics & Biochemistry, Yale University, New Haven, CT

Research Experience

2008–2013: Graduate Studies
    Laboratory of Prof. David Bartel, Massachusetts Institute of Technology, Whitehead Institute
    Discovery and biochemical characterization of RNA interference in budding yeast

2004–2007: Undergraduate Studies
    Laboratory of Prof. Sandra Wolin, Yale University
    Role of noncoding Y RNAs in the cellular function of the Ro 60-kDa autoantigen

Honors and Awards

2013: Harold M. Weintraub Graduate Student Award, Fred Hutchinson Cancer Research Center

2012: UCSF Faculty Fellow offer, University of California, San Francisco

2011: Abraham J. Siegel Fellowship Award, Whitehead Institute for Biomedical Research

2011: EMBO Prize for best PhD student talk, 6th Microsymposium on Small RNAs

2007–2010: Graduate Research Fellowship, National Science Foundation

2007: *Summa cum laude* with Distinction in Molecular Biophysics & Biochemistry, Yale University

2007: Paul Sigler Memorial Prize, Yale University

2006: Goldwater Scholarship, Barry M. Goldwater Scholarship and Excellence in Education Foundation

2005: Phi Beta Kappa, Yale University

Publications

Nakanishi K*, **Weinberg DE***, Bartel DP, Patel DJ (2012). Structure of yeast Argonaute with guide RNA. *Nature* 486:368–74.

Bernstein DA*, Vyas VK*, **Weinberg DE**, Drinnenberg IA, Grisafi PG, Bartel DP, Fink GR (2012). *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc Natl Acad Sci USA* 109:523–8.

Sim S, Yao J, **Weinberg DE**, Niessen S, Yates III JR, Wolin SL (2012). The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-kDa autoantigen and the nonconding Y3 RNA. *RNA* 18:100–10.

Weinberg DE*, Nakanishi K*, Patel DJ, Bartel DP (2011). The inside-out mechanism of Dicers from budding yeasts. *Cell* 146:262–76.

Drinnenberg IA*, **Weinberg DE***, Xie KT*, Mower JP, Wolfe KH, Fink GR, Bartel DP (2009). RNAi in budding yeast. *Science* 326:544–50.

Sim S, **Weinberg DE**, Fuchs G, Choi K, Chung J, Wolin SL (2009). The subcellular distribution of an RNA quality control protein, the Ro autoantigen, is regulated by noncoding Y RNA binding. *Mol Biol Cell* 205:1555–64.

** indicates equal contributions*

Patents

Bartel DP, Drinnenberg IA, **Weinberg DE**, Xie KT, Wolfe KH, Fink GR. RNAi in Budding Yeast. United States Patent Application Number 13/394,834. Filing Date: September 10, 2010.

Oral Presentations

Faculty Fellow Applicant Seminars (Harvard University FAS Center for Systems Biology; University of California, San Francisco; Cold Spring Harbor Laboratory), November 2012. "Insights into post-transcriptional gene regulation: RNAi and the translating mRNP"

EMBO | EMBL Symposium on the Complex Life of mRNA, October 2012, Heidelberg, Germany. "The structure of yeast Argonaute with guide RNA"

FASEB Science Research Conference on Nucleic Acid Enzymes, June 2012, Snowmass Village, CO, USA. "The structure of yeast Argonaute with guide RNA"

The Gordon Research Conference on Nucleic Acids, June 2011, Biddeford, ME, USA. "The inside-out mechanism of Dicers from budding yeasts"

6th Microsymposium on Small RNAs, May 2011, Vienna, Austria. "The inside-out mechanism of Dicers from budding yeasts"

15th Annual Meeting of the RNA Society, June 2010, Seattle, WA, USA. "The inside-out mechanism of Dicers from budding yeasts"

New England RNA Data Club Symposium at New England Biolabs, May 2010, Ipswich, MA, USA. "The inside-out mechanism of Dicers from budding yeasts"

Teaching Experience

Spring 2011: Teaching Assistant, General Biochemistry (7.05), MIT

Fall 2008: Teaching Assistant, Principles of Biochemical Analysis (7.51), MIT

Fall 2006–Spring 2007: Undergraduate Teaching Intern, Freshman Organic Chemistry, Yale University

Fall 2005–Spring 2007: Biochemistry Tutor, Science Tutoring Program, Yale University